DOCUMENT RESUME

ED 429 391

EC 307 129

AUTHOR	Callahan, Carolyn M.; Tomlinson, Carol A.; Hunsaker, Scott L.; Bland, Lori C.; Moon, Tonya
TITLE	Instruments and Evaluation Designs Used in Gifted Programs.
INSTITUTION	National Research Center on the Gifted and Talented, Storrs, CT.; Virginia Univ., Charlottesville.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
REPORT NO	RM-95132
PUB DATE	1995-09-00
NOTE	149p.
CONTRACT	R206R00001
AVAILABLE FROM	NRC/GT, University of Connecticut, 362 Fairfield Road, U-7, Storrs, CT 06269-2007.
PUB TYPE	Reports - Research (143)
EDRS PRICE	MF01/PC06 Plus Postage.
DESCRIPTORS	Case Studies; *Data Analysis; Data Collection; Databases; Decision Making; Educational Practices; Elementary Secondary Education; *Evaluation Criteria; *Evaluation Methods; *Gifted; National Surveys; Program Development; *Program Evaluation; Research Design; Research Methodology

ABSTRACT

This report presents results of a series of studies documenting current practices in the evaluation of gifted programs and investigating the factors which make evaluation more useful to decision-makers. The investigation involved establishing several databases containing three kinds of information: (1) abstracts of articles relating to evaluation utility and the evaluation of gifted programs; (2) instruments that have been used by other school districts in the evaluation of gifted programs as well as reviews of these instruments; and (3) actual evaluations used across the nation to assess the effectiveness of gifted programs. Studies identified factors which improve the likelihood that results of an evaluation will be useful and will lead to development of a set of guidelines. Chapter 1 provides an introduction to the National Repository databases. Chapter 2 reviews the literature on program evaluation. Chapter 3 reports on a study of current practices in the evaluation of gifted programs. Chapter 4 presents case studies in program evaluation utilization in gifted programs. Chapter 5 offers a summary and conclusions. Ten appendices include tables, the scale developed for evaluation of the program evaluation instruments, a planning guide for program evaluation, the program evaluation guidelines, and the evaluation instruments database form. (Contains 55 references.) (DB)

****

* Reproductions supplied by EDRS are the best that can be made *

* from the original document. *

****

ED 429 391

NRC
G/T

**THE NATIONAL
RESEARCH CENTER
ON THE GIFTED
AND TALENTED**

The University of Connecticut
The University of Georgia
The University of Virginia
Yale University

Instruments and Evaluation Designs Used in Gifted Programs

Carolyn M. Callahan
Carol A. Tomlinson
Scott L. Hunsaker
Lori C. Bland
Tonya Moon

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it
☐ Minor changes have been made to improve reproduction quality
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

The University of Virginia
Charlottesville, Virginia

September 1995
Research Monograph 95132

EC 307129

BEST COPY AVAILABLE

Instruments and Evaluation Designs Used in Gifted Programs

Carolyn M. Callahan
Carol A. Tomlinson
Scott L. Hunsaker
Lori C. Bland
Tonya Moon

The University of Virginia
Charlottesville, Virginia

September 1995
Research Monograph 95132

THE NATIONAL RESEARCH CENTER ON THE GIFTED AND TALENTED

The National Research Center on the Gifted and Talented (NRC/GT) is funded under the Jacob K. Javits Gifted and Talented Students Education Act, Office of Educational Research and Improvement, United States Department of Education.

The Directorate of the NRC/GT serves as the administrative unit and is located at The University of Connecticut.

The participating universities include The University of Georgia, The University of Virginia, and Yale University, as well as a research unit at The University of Connecticut.

The University of Connecticut
Dr. Joseph S. Renzulli, Director
Dr. E. Jean Gubbins, Assistant Director

The University of Connecticut
Dr. Francis X. Archambault, Associate Director

The University of Georgia
Dr. Mary M. Frasier, Associate Director

The University of Virginia
Dr. Carolyn M. Callahan, Associate Director

Yale University
Dr. Robert J. Sternberg, Associate Director

Copies of this report are available from:
NRC/GT
The University of Connecticut
362 Fairfield Road, U-7
Storrs, CT 06269-2007

Research for this report was supported under the Javits Act Program (Grant No. R206R00001) as administered by the Office of Educational Research and Improvement, U.S. Department of Education. Grantees undertaking such projects are encouraged to express freely their professional judgement. This report, therefore, does not necessarily represent positions or policies of the Government, and no official endorsement should be inferred.

Note to Readers...

All papers by The National Research Center on the Gifted and Talented may be reproduced in their entirety or in sections. All reproductions, whether in part or whole, should include the following statement:

> **Research for this report was supported under the Javits Act Program (Grant No. R206R00001) as administered by the Office of Educational Research and Improvement, U.S. Department of Education. Grantees undertaking such projects are encouraged to express freely their professional judgement. This report, therefore, does not necessarily represent positions or policies of the Government, and no official endorsement should be inferred.**
>
> **This document has been reproduced with the permission of The National Research Center on the Gifted and Talented.**

If sections of the papers are printed in other publications, please forward a copy to:

The National Research Center on the Gifted and Talented
The University of Connecticut
362 Fairfield Road, U-7
Storrs, CT 06269-2007

Please Note: Papers may not be reproduced by means of electronic media.

Acknowledgements

Many individuals have contributed to this research effort and we thank them for their help. We appreciate help in test and file review from A. Christian Lundberg, Paula Pizzat, Ari Rapkin, Caroline Cunningham, Marcia Delcourt, Lisa Markel, Catherine Brighton, Laurie Abeel, Kim Mitchell, and Karen Lelli. Additional help with data entry and management came from Lori Lutz, S. Christopher Roberson, Angela Roach, and Thu Vu. Ann Holthaus, Anne Gwathmey, Ruth Lyman, and Wanda Lam provided invaluable clerical support. John Wesley Taylor provided valuable editing assistance in the production of the final report. We thank the teachers, administrators, and researchers who contributed material to the data bases and participated in the research studies. We also thank other university faculty who were involved at various points in the project, particularly Michael S. Caldwell.

Instruments and Evaluation Designs Used in Gifted Programs[*]

Carolyn M. Callahan
Carol A. Tomlinson
Scott L. Hunsaker
Lori C. Bland
Tonya Moon

The University of Virginia
Charlottesville, Virginia

ABSTRACT

The project entitled *Investigations Into Instruments Used in the Identification of Gifted Students and the Evaluation of Gifted Programs* was divided into two avenues of study. The first series of inquiries are reported in the technical report document entitled *Instruments Used in the Identification of Gifted and Talented Students*. The second series of studies, which are reported in this document, focused on documenting current practices in the evaluation of gifted programs and on investigating the factors which make evaluation more useful to decision-makers. A solicitation of instruments and program evaluation designs led to the establishment of databases containing information on current practices. The review of current practices and a study of evaluation utility provided us with guidelines for constructing useful and informative evaluations, some disappointing findings which indicate that often these guidelines are not being followed, and heartening examples of promising practices.

---

[*] This report is the second of two technical reports which summarize the research project entitled "Investigations Into Instruments Used in the Identification of Gifted Students and the Evaluation of Gifted Programs."

Instruments and Evaluation Designs Used in Gifted Programs

Carolyn M. Callahan
Carol A. Tomlinson
Scott L. Hunsaker
Lori C. Bland
Tonya Moon

The University of Virginia
Charlottesville, Virginia

EXECUTIVE SUMMARY

The ability of persons providing services to gifted children to improve those services and increase effectiveness and efficiency of programming efforts is dependent on having reliable and valid information about the current status of the program and the outcomes that are being achieved by the program. However, the literature in gifted education has repeatedly asserted that little program evaluation occurs in this field and that the evaluations that are conducted are not adequate to provide the needed types of information. Within the context of this study *Instruments and Evaluation Designs Used in Gifted Programs*, we have explored the validity of the assertions made in the literature about evaluations of programs for the gifted, analyzed current evaluation literature for generic guidelines for effective evaluations, and studied the utility of evaluations of programs for the gifted with the intent of providing more specific guidelines for decision makers in the construction of evaluation designs, implementation of the evaluation process, and utilization of evaluation results.

As the first step in providing both professional evaluators and school practitioners with tools to use in the evaluation process, we compiled several databases containing three kinds of information. The first set of databases contains abstracts of articles relating to evaluation utility and the evaluation of gifted programs. The second set of databases is comprised of instruments that have been used by other school districts in the evaluation of gifted programs as well as reviews of these instruments on an instrument developed in accordance with standards for instrument design and use provided by various professional associations such as the National Council on Measurement in Education. Finally, we have a collection of actual evaluations used across the nation to assess the effectiveness of gifted programs. These databases are accessible through contacting the University of Virginia site of The National Research Center on the Gifted and Talented.

From the literature in the field of evaluation we extracted a series of factors which improve the likelihood that the results of any evaluation will be useful:

1. Begin with adequate funds and staff commitments. This should include both funds for carrying out the evaluation and funds for implementing change. Evaluation results are less likely to be used, or to be used appropriately, if there are no funds to implement recommendations. Lack of commitment to the program, or to program change on the part of people in positions of power and influence, results in little attention to evaluation findings.
2. Select clear, appropriate designs:

    a. It is most effective to develop evaluation plans at the earliest stages of program planning, the earlier program evaluation planning occurs the more likely the evaluation will be used to help form good services.
    b. It is also critical to define the purposes of the evaluation, and to select an evaluation design appropriate to the program and program features that will be evaluated. For example, quantitative designs may be especially useful when outcomes are a focus; however, qualitative designs are more appropriate when processes within a program are studied or when complex settings are examined. A combination of qualitative and quantitative designs is called for when both processes and outcomes are of concern.
3. Establish credibility of evaluator and evaluation process. It is important that the evaluator be respected by those who will receive the evaluation report.
4. The evaluator should carefully explain procedures and rationales at the outset of the process and then again at the time of the report so that the consumers will clearly understand the strategies used in determining findings and recommendations.
5. Information collected should be of sufficient breadth and collected in ways which allow pertinent questions (questions of significance to the decision makers) to be pursued. The data should be collected and analyzed in ways which address the needs of a variety of appropriate, interested audiences.
6. Use multiple data gathering methods (e.g., surveys, observations, interviews). Standardized measures increase the usefulness of findings and draw upon a variety of data sources (students, teachers, parents, school board members, administrators).
7. Prepare understandable, timely and well-documented, but succinct reports. Similarly, it is important that the evaluation report be disseminated to clients and relevant audiences in a timely fashion, which allows information to be received while it is useful and can be acted upon.
8. Direct reports to appropriate audiences at appropriate times. It is important to clearly identify clients and audiences of the evaluation, and to involve them actively throughout the evaluation design, data collection, and data analysis. People who feel a clear need for evaluation are more likely to utilize findings than those who do not.
9. Maintain effective and on-going communication with clients to establish the worth of the evaluation.

The literature also identified specific challenges facing an evaluator of programs for gifted students. Suggestions which emerged for dealing with issues relating to design or articulation of the programs themselves and issues of evaluation design and measurement include:

1. Clearly delineate program goals—both long term and short term in clearly understood terms which can be operationally defined. Programs for the gifted often suffer from poorly delineated program goals. In instances where program goals are unstated, vague or unfocused, it is difficult to design an evaluation that addresses the impact of the program.
2. Carefully address design and measurement issues. Many of the confounding traits of programs for gifted learners have an impact on measurement and design decisions within the evaluation. Carefully assure that the instruments selected for assessing program goals are valid and reliable, do not suffer from ceiling effects. Allow for control of regression to the mean effects:

a. Use out-of-level tests where *valid* for the trait/outcome assessed to combat the low ceiling effect.

b. Develop and use common, valid criteria for examining student products and portfolios, and establish inter-rater reliability in application of the criteria.

3. While the use of control groups is difficult, stakeholders may require some evidence that program effects are a result of services, not maturation. As alternatives to randomized experiments, consider the use of carefully matched groups between schools; a time-series design in which all groups of gifted learners receive the target intervention, but at various times, thus serving as controls for one another; contrast group (rather than a control group) in which an existing group or to-be-generated data set serves as a contrast to results from the intervention in question.
4. Prepare staff carefully for the evaluation. Ensure that both staff and evaluators are trained to carry out and analyze the results of the evaluation. Prepare staff and describe rules of scoring prior to administration of tests.
5. Address questions important to the evaluation audiences. Address the needs of both internal and external audiences of programs, and address questions helpful in making decisions which can have an impact on program quality. Consider goals, activities, and structure of the program in question; questions relating to program areas which are of central importance or present potential problems in the program; questions relating to level of resources, undesirable change brought about by the program, conflict with values of other stakeholders, loss of power, inconsistency between program goals and implementation of those goals, lack of understanding of goals, and personal bias.
6. Evaluation questions should be specific to the program being evaluated, unlike research questions which seek generalizability to other settings.
7. Use a variety of data collection strategies.
8. Know the biases of decision-makers. In regard to characteristics which may affect utilization of evaluation results in programs for the gifted, it is necessary for the evaluator to identify decision-makers clearly and to understand the actions over which they have control. Find out what courses of action will result from evaluation findings, and make recommendations with an eye toward improving the program.

Using the results of these literature reviews and the results of the evaluation utility study, a set of "Guidelines for Evaluating Gifted Programs" was prepared.

Preparing for the Evaluation

Much of the success of a program evaluation will depend on the quality of decisions made prior to actually conducting the evaluation. Planning is an essential phase of the process and should proceed carefully and thoughtfully.

- Does the program have clearly articulated goals and objectives which can be a focus of evaluation?
- Are the articulated goals and objectives the ones valued as a program focus?
- Does the school district have a commitment to meaningful evaluation of programs including adequate time, finances, and personnel time given to evaluation and dissemination of findings?

- Have you identified representatives of varied internal and external interest groups or stakeholders (e.g., parents, regular classroom teachers, administrators, students, gifted/talented specialists, school board members, representatives of business and industry) to serve as an active evaluation steering committee which will be involved in setting the parameters of the evaluation?
- Is there a written plan for evaluating the program, including delineated steps and procedures in the process?
- Is there a plan for on-going feedback during the evaluation (formative as well as summative evaluation)?
- Are the evaluators knowledgeable about both gifted education and evaluation?
- Are the evaluators knowledgeable about both qualitative and quantitative research strategies?
- Do evaluators, program personnel and/or steering committee members include those with sufficient political sophistication to understand the political implications of evaluation? Can they aid in identifying and gaining access to key decision makers and can they provide an understanding of the actions over which the decision-makers have control?
- Are roles of evaluators, administrators, stakeholders, and steering committee members in the evaluation process clearly articulated?
- Is there a working plan to develop networks of support both inside and outside the school district for the evaluation process, its findings, and the program?
- Are there appropriate time lines for data gathering, analysis, and dissemination?
- Will the evaluation data be collected, analyzed, and presented in time to influence decision-making?
- Are there plans for monitoring processes and procedures throughout the evaluation?
- Are appropriate provisions established to ensure confidentiality and sensitivity in handling data?
- Are there clearly stated evaluation questions that appropriately address program goals, structures, functions, and/or activities?
- Do the evaluation questions seem likely to generate findings that will have a positive impact on programs and participants?
- Are there plans to use multiple data sources (e.g., parents, regular classroom teachers, identified students, other students, gifted education specialists, administrators) in order to understand perspectives of various stakeholders?
- Are there plans to employ varied data collection modes (e.g., face-to-face interviews, telephone interviews, classroom observations, group meetings, product reviews, staff development evaluations, mail out surveys, test data) in order to reflect the complex nature of the program and meet data needs of various constituencies?
- Do potential users of findings have opportunities to provide input on types of information desired and forms in which the information would be most usefully reported?
- Have you examined ways to collect "process data" which can show whether the program is functioning as it should?
- Have you examined way to collect "outcome data" which can show whether student affective and/or academic growth has occurred as a result of program participation?
- Have you considered ways in which case study data can be useful to document program effectiveness?

- Have you selected reliable and valid assessment tools?
- Have you described ways in which data will be analyzed?
- Have you specified ways in which data will be reported to various groups?
- Have you prepared staff members for the data-collection phase of the evaluation process and their roles in it?
- Are multiple stakeholders consistently involved with data collection?
- Are program evaluators consistently visible to varied audiences to facilitate understanding of those audiences by the evaluators and understanding of the program and evaluation process by the audiences?
- Are multiple stakeholders consistently involved with monitoring and reviewing the evaluation process and its evolving findings?
- Do you have a plan for quick turnaround time for data analysis and feedback, with specific guidelines for all individuals in meeting prescribed time lines?
- Is there a commitment from evaluators, key program personnel, and steering committee members to the use of findings for positive program change?
- Is there an articulated plan for turning findings into action, incorporating the roles which evaluators, program personnel, and stakeholders will play in that process?
- Have evaluators, program personnel, and evaluators assessed the impact of evaluation findings?
- Are findings prepared and interpreted according to interest and needs of stakeholder groups?
- Are evaluation reports clear? Do they avoid the use of jargon and confusing technical interpretations of data?
- Do evaluation reports describe the program, evaluation questions, evaluation process, participants in the process, data collection, and data analysis?
- Are evaluation reports designed for follow-through with specific recommendations made for acting upon findings?
- Are evaluation reports and recommendations presented to decision-makers in a timely fashion?
- Are there provisions for oral explanations and discussions of findings with stakeholders and decision-makers?
- Has the steering committee assessed the evaluation process according to initial goals, roles and time lines, including making written recommendations for changes in the next evaluation cycle?
- Have evaluators, steering committee members, and program personnel followed up with policy makers until appropriate actions have been taken?
- Has the steering committee proposed questions for further examination in upcoming evaluation cycles and resulting from insights gained in the current evaluation cycle?

Table of Contents

ABSTRACT vii

EXECUTIVE SUMMARY ix

CHAPTER 1: Introduction to the National Repository 1
- Report Overview 2
- Establishment of the National Repository 2
  - Mailing 2
  - Announcements 3
  - Responses 3
  - Review of the Literature 4
  - Establishing Databases 4
- Data Analysis 4
- Analysis of Evaluation Instruments Related to Evaluation Questions or Outcomes 7
- Assessing the Psychometric Properties of Published Instruments 8
  - Validity and Reliability of *SEPEI* 9
    - Content Validity 9
    - Reliability 9
  - Articles, Test Reviews, and Locally Developed Instruments 10
- Importance of This Repository 10

CHAPTER 2: Literature Review 11
- Setting the Stage for Focusing on Evaluation 11
- Evaluation Utilization Rationale 11
- Data Gathering 12
- Definitions 12
- Purposes for Evaluating 13
- Evaluation Design 13
- Factors Affecting the Use of Evaluation Findings 14
  - Evaluation Context Factors 14
  - Evaluator Control Factors 14
    - Message Source 14
    - Message Content 15
    - Characteristics of the Receiver 15
    - Communication in General 15
    - Use of Multiple Data Collection and Reporting Strategies 15
- Utilization of Evaluation Results in Gifted and Talented Programs 16
  - Special Challenges in Evaluating Gifted Programs 16
- Recommended Evaluation Designs 17
- Utility Standards and Evaluation of Gifted and Talented Programs 18
  - Recommendations Related to Message Source 19
  - Recommendations Related to Message Content 19
  - Recommendations Related to Receiver Characteristics and Audience Identification 19
  - Applied Evaluation in Gifted and Talented Programs 20

Table of Contents (continued)

- **CHAPTER 3: Current Practices in the Evaluation of Gifted Programs** 21
  - Methodology 21
    - Data Gathering 21
    - Data Analysis 22
    - Limitations 22
  - Results 22
    - A Typical Evaluation 22
    - Specific Details of the Frequency Analysis 23
      - Evaluation Types 23
      - Evaluation Models 23
      - Evaluator Types 23
      - Data Gathering Methods 23
      - Data Analysis Techniques 24
      - Data Sources 24
      - Intended Audiences 24
      - Evaluation Concerns 24
      - Reporting Formats 24
      - Utility Practices 24
  - Specific Chi Square Analyses 25
    - Concerns and Promising Practices 26

- **CHAPTER 4: Case Studies in Program Evaluation Utilization in Gifted Programs** 27
  - Rationale 27
  - Method 27
    - Definition 27
    - Data Gathering 27
    - Data Analysis 28
  - Results 29
    - Looking at the Group Characteristics 29
  - The Continua Described 30
    - Evaluation Focus 30
    - Participants in the Process 30
    - Methods of Evaluation and Data Analysis 31
    - Implementation Plans 31
    - Evaluation Reports 31
    - Purposes for Evaluations 32
    - Qualifications of Program Personnel 32
    - Nature of Change Resulting From the Evaluation Process 33
    - Profiles of the Districts 33
    - Profile of the Evaluation Process in a District With an Evaluation Characterized as Weaker 34
    - Profile of the Evaluation Process in a District With an Evaluation Characterized as Stronger 34
    - A Cross-Group Comparison 35
    - A Tale From the Middle Group 36
    - Decisive Factors in Use of Findings 37
  - Concerns 37
  - Promising Practices 38

Table of Contents (continued)

CHAPTER 5: Conclusions and Summary 39
- Increasing Use of Evaluation Results in General: The Literature 39
  - The Impact of Economics, Politics, Definition, and Design 39
- The Impact of Message Source, Content, and Receiver 40
- Special Challenges in Evaluating Gifted Programs 40
- The Messenger, the Receiver, and the Evaluations for Gifted Programs 42
- Summary Guidelines for Conducting Useful Evaluations of Gifted Programs 43
- Increasing Evaluation Utilization: Our Studies 43
- Suggestions for Improving Gifted Program Evaluation 44

References 47

Appendices

- Appendix A: Tables 51
- Appendix B: *Scale for the Evaluation of Program Evaluation Instruments* (*SEPEI*) 65
- Appendix C: Technical Manual for the *Scale for the Evaluation of Program Evaluation Instruments* (*SEPEI*) 71
- Appendix D: *SEPEI* Inter-rater Item Descriptive Statistics for the *Cornell Test of Critical Thinking* and the *Ross Test of Higher Cognitive Processes* 81
- Appendix E: A Typical Response to Request for Data from the Evaluation Database 87
- Appendix F: A Planning Guide for Evaluating Programs for Gifted Learners 107
- Appendix G: Guidelines for Conducting Useful Evaluations of Programs for Gifted Learners 113
- Appendix H: Reference on Evaluation Utility Classified According to the Factors Established by the Joint Committee on the Standards for Educational Evaluation 119
- Appendix I: Second Round of Interview Questions 123
- Appendix J: Evaluation Instruments Database Form 127

List of Figures

Figure 1. Standardized Instruments Used to Assess Program Evaluation Questions 5

Figure 2. List of Standardized Instruments Used but Unrelated to a Specific Evaluation Question 7

Instruments and Evaluation Designs Used in Gifted Programs

Carolyn M. Callahan
Carol A. Tomlinson
Scott L. Hunsaker
Lori C. Bland
Tonya Moon

The University of Virginia
Charlottesville, Virginia

CHAPTER 1: Introduction to the National Repository

The goals of the Identification and Evaluation Project conducted by the NRC/GT at the University of Virginia were (a) to identify current practices in identifying gifted students and in evaluating gifted programs; (b) to collect relevant data on assessment instruments, (c) to evaluate those instruments using standards established by the measurement field, and (d) to identify promising practices in identification and evaluation. The first stage of the project was the establishment of a National Repository for Instruments and Strategies Used in the Identification and Evaluation of Gifted Students Programs. The second phase involved reviewing available data, including reliability and validity data, on identification and evaluation instruments in the Repository and rating the instruments on their appropriateness for specific purposes. During the third phase we investigated the effectiveness of promising non-published identification instruments. Studies of identification instruments and procedures were conducted concurrently and are described in separate publication, *Instruments Used in the Identification of Gifted and Talented Students.* Finally, promising innovative practices for identifying students from at-risk populations were identified from entries in our data bank and the model projects funded through the Jacob K. Javits Gifted and Talented Education Act program. Descriptions of these practices were compiled into a separate monograph, *Contexts for Promise: Noteworthy Practices in the Identification of Gifted Students* (Callahan, Tomlinson, & Pizzat, n.d.).

The initial focus of our investigation emphasized collecting and evaluating extant identification and evaluation literature, instruments, systems, and designs. The major research questions posed for the identification aspect of the study included: What are the most commonly used instruments in identifying gifted and talented students? What instruments are used for identifying gifted and talented students according to specific definitions and conceptions of giftedness? What evidence is there of the reliability and validity of these instruments, and is that evidence sufficient to justify their use with given definitions of giftedness and for identifying underserved populations?

Similar questions were posed regarding evaluation instruments and designs: What instruments are most commonly used in the evaluation of gifted students and programs? What are the reliability and validity of these instruments in assessing goals and objectives common to gifted programs? What instruments (especially non-traditional and product-oriented instruments) are used to evaluate programs for the gifted and talented? Which evaluation designs or which characteristics of evaluation designs yield useful evidence in program development and modification?

During the second stage of this investigation, three non-published specific instruments potentially useful in identifying underserved gifted students were selected for further investigation of their psychometric properties. The major research questions in this stage of the study were: What are the reliability and validity of each of these instruments? How effective are these instruments in identifying underserved populations of gifted students? In each case, we investigated the effectiveness of instruments relative to particular definitions of giftedness or the particular stated outcome goals of gifted programs.

In preparing the monograph on promising practices, the following questions were used as guides: Are there systems with documented evidence of effectiveness for identifying the underserved gifted? Do these systems used in identifying typically underserved gifted and talented students result in the identification of students who have special talents and needs?

The first evaluation study focused on an analysis of frequency of type of evaluation (summative/formative), evaluation model (management centered, objective centered, etc.), evaluator type (external/internal), data-gathering methodology, data analysis technique, data sources, audiences, evaluation concerns, report formats, and recommendations.

Report Overview

Because different portions of the project had different methodologies, each chapter of this report centers on one aspect of the study. This chapter presents the establishment of the National Repository. Chapter 2 presents the review of current literature on the evaluation of gifted programs. Chapter 3 summarizes the results of analyzing the characteristics of reports submitted to the Repository. The findings of the evaluation utilization study are presented in Chapter 4.

Establishment of the National Repository

Mailing

To gather as many instruments, identification strategies, and evaluation designs as possible, we designed a process to gather information on both standardized and locally developed instruments, and to identify state and local evaluation designs. Specific efforts were made to identify instruments and strategies which had been used in the identification of minority, economically disadvantaged, non-English speaking, and handicapped gifted students, and in evaluating programs for these students.

Four strategies were employed to collect the instruments, systems, and designs that have been used for program evaluation and student identification at the national, state, regional, and local levels. First, a letter requesting all state criteria used in identification systems, state recommended identification instruments, state-wide evaluation reports, and evaluation instruments was sent to each official in the state departments of education, who had been designated (as of Fall, 1990) as having responsibility for gifted and talented programs. These individuals were asked to supply copies of any identification or evaluation instruments being used on a state, regional, and/or local level or to provide a list of district level personnel who could be contacted for such information. They were asked to furnish the name of the developer of the instrument, information on how the instrument was used, who used it (i.e., psychologist, teacher, evaluator), and how data were analyzed. State officials were advised that they could submit state guidelines, evaluation reports, or other documents from which we would glean the necessary information if that were more convenient.

Next, each Collaborative School District (a CSD is a school district that had specifically agreed to work on NRC/GT projects) was asked (through a mailing) to provide any instruments used in identifying gifted students, a description of identification procedures used, demographic information on students selected, and copies of any evaluations of their programs or projects in gifted education. They were also asked the name of the instrument developer and the uses to which the instrument was put. We also asked that, whenever possible, the name of the evaluator be provided.

A similar letter and form were sent to approximately 5,000 school districts across the United States. Addresses for these districts were obtained from an educational database firm. Where possible, we delivered the letters at state conferences (Florida, Iowa, and Virginia), through state association mailings (Texas), and through state gifted coordinators (Colorado and Arizona).

We recognized, of course, that districts might not be comfortable with their current identification procedures or instruments, or districts might realize that they didn't truly abide by stated procedures or state regulations, and therefore, might be reluctant to respond accurately (or at all) to the survey. We attempted to avoid any bias that might arise in the responses in two ways. First, districts were assured that information would be strictly confidential and we would not reveal names of districts in our reporting of data without the school district's permission. Second, our survey clearly emphasized that we were interested in all data about instruments and surveys, including instruments or systems which didn't seem to work as intended. We stressed the importance of learning from the things that do not function as expected, as well as learning from the things that do work. Requests concerning the value of each instrument sought respondents' information on the positive and negative aspects of the instruments in general, as well as information on identifying students from specific underserved populations. Finally, a random sample of non-respondents was contacted by follow-up letter to determine whether there had been a systematic response bias.

All contacts were asked specifically to indicate instruments, strategies, and data sources that they believed had been particularly useful in identifying minority, economically disadvantaged, underachieving, non-English speaking, and/or handicapped gifted students. The Council for Exceptional Children and state department personnel were asked for lists of institutions that specifically serve individuals who are blind, or hearing impaired, or have other handicapping conditions so that they could be contacted specifically and directly. In addition, all individuals contacted were asked for program evaluation instruments, including process and product/performance ratings, and standardized tests.

Announcements

Professional organizations, journals, and state associations through which it would be appropriate to make requests for information were identified and specifically tailored announcements and letters were sent to each association and journal. In addition, announcements were included in the conference programs and/or registration packets at the annual meetings of the National Association for Gifted Children and the American Evaluation Association.

Responses

The mailings and announcements yielded responses containing identification or evaluation information from 542 individual school districts. An additional 65 school districts responded that they would have liked to forward materials, but could not do so because the program had recently been cut or was undergoing extensive changes. A random sample of 140 non-responding CSDs and 100 additional non-responding (local education agencies [LEAs], but not CSDs) was sent a questionnaire asking why they had not responded. Of these, 45 CSDs and 44 LEAs returned the questionnaire. Results of that survey are reported in *Instruments Used in the Identification of Gifted and Talented Students.*

Review of the Literature

Database searches were conducted across Educational Resources Information Center (ERIC), PsycLIT (the computerized version of *Psychological Index), Dissertation Abstracts International,* and VIRGO *(the University of Virginia computerized card catalogue system).* Search terms included gifted, ratings, scales, reliability, validity, tests, measurements, evaluations, and utilization. These terms were used singly or in combination as appropriate. Each search yielded a list of potential resources which were reviewed for information on the state of the art in identification or evaluation (particularly evaluation utilization), information on use of particular instruments or strategies for identification or evaluation, and information on reliability and validity.

The initial search yielded 375 documents on identification and/or program evaluation in gifted education including approximately 174 journal articles, 16 books, 37 dissertations, and 120 ERIC documents. In some cases dissertations were obtained directly from the authors. Large ERIC documents were reviewed on microfiche with copies made of relevant sections only. Abstracts of each document were prepared focusing particularly on either test review information or usefulness in identifying underserved gifted students.

Establishing Databases

The information compiled from the resources listed above yielded four databases on evaluation as part of the National Repository. The computer databases cover three categories of information: bibliographic entries, standardized instrument reviews and use, and locally developed materials. The bibliographic databases contain abstracts of published reviews of standardized instruments, abstracts of articles about the use of standardized instruments in evaluation, and abstracts of articles about particular issues in evaluation (e.g., underserved populations). The standardized instrument databases include listings of the ways in which published instruments are used and reviews of the instruments on NRC/GT developed scales. The local instrument databases include listings of a collection of identification instruments developed and used at the local school level but not published. Within each database, the entries are further divided into two groups—those we have permission to share with the public and those we do not. A complete list of the evaluation database names, content descriptions, and number of entries appears as Table A-1 (see Appendix A). The particular categories were created in order to facilitate searches for information by project staff and ultimately by educators, psychologists, and parents seeking information from the databases. While a particular article might relate to more than one category, it was classified by the dominant theme of the article.

Data Analysis

For each evaluation report received, we identified questions or goals of the evaluation as listed or implied in the report. From each report, we determined which standardized instruments addressed which evaluation question. The evaluation questions/goals were grouped into these outcome categories: achievement, aptitude, attitudes toward others, autonomy/responsibility, creativity, general academic outcomes, general affective outcomes, general program outcomes or effectiveness, general student growth, identification, locus of control, research skills, self-concept, student perceptions of school/program, study habits, and thinking skills. Each standardized instrument used in each report was catalogued into the appropriate evaluation questions category. Then we counted how often each standardized instrument was used to evaluate a given outcome of the program evaluation questions or goals (Figure 1).

Category of Evaluation Question	Name of Instrument	Number of Reports Citing Use
Achievement	California Achievement Test	2
	Clymer Barrett	1
	Comprehensive Test of Basic Skills	5
	Iowa Tests of Basic Skills	1
	Metropolitan Achievement Test	1
	Peabody Picture Vocabulary Test	1
	Preliminary Scholastic Aptitude Test	1
	Scholastic Aptitude Test	1
	Sequential Tests of Educational Progress, Series III	1
	Stanford Achievement Test	3
	Texas Educational Assessment of Minimal Skills	1
	Test of Academic Aptitude	1
Aptitude	Developing Cognitive Abilities Test	1
Attitudes toward others	School Situation Survey	1
Autonomy/ Responsibility	Intellectual Achievement Responsibility Scale	1
Creativity	Something About Myself	1
	Student Product Assessment Form	1
	Test of Creative Potential	1
	Thinking Creatively in Action and Movement	1
	Torrance Tests of Creative Thinking	2
	Torrance Tests of Creative Thinking—Demonstrator Form	1
	Wallach-Kogan Creativity Instrument	2
General Affective Outcomes	Dimensions of Self Concept Inventory	1
General Program Outcomes	California Achievement Test	2
	Clymer Barrett	1
	Comprehensive Test of Basic Skills	1
	Criterion Referenced Talent Tests	1
	Preliminary Scholastic Aptitude Test	1
	Stanford Achievement Test	1
	Torrance Tests of Creative Thinking	1

Figure 1. Standardized instruments used to assess program evaluation questions.

(figure continues)

Category of Evaluation Question	Name of Instrument	Number of Reports Citing Use
General Student Growth	California Achievement Test	1
	Cornell Critical Thinking	1
	Ross Test of High Cognitive Processes	2
	Torrance Test of Creative Thinking	1
Identification	California Achievement Test	3
	Cognitive Abilities Test	1
	Comprehensive Test of Basic Skills	1
	Culture Free Self Esteem Inventory	1
	Iowa Tests of Basic Skills	2
	Matrix Analogies Test	1
	Otis-Lennon Mental Abilities Test	1
	Otis-Lennon School Abilities Test	1
	Scales for Rating the Behavioral Characteristics of Superior Students	1
	Scholastic Aptitude Test	1
	Stanford Achievement Test	1
	Stanford-Binet	1
	Structure of Intellect Gifted Screening Form	1
	Test of Divergent Thinking	1
	Test of Cognitive Skills	1
	Wechsler Intelligence Scale for Children-Revised	2
Locus of Control	James' Internal/External Locus of Control	1
Research Skills	GAIN Teacher Assessment of Student Research Skills	1
Self-concept	Coopersmith Test of Self-Esteem	1
	Charter Self-Perception Profile	1
	ME Scale	3
	Piers Harris Children's Self Concept Scale	1
	Revised Janis-Field Feeling of Inadequacy Scale	1
	Self-perception Inventory	1
Student Perceptions	Quality of School Life	2
Study Habits	Survey of Study Habits and Attitudes	1
Thinking Skills	Cognitive Ability Test	1
	Criterion Referenced Talent Tests	2
	Developing Cognitive Abilities Test	2
	Ross Test of High Cognitive Processes	8
	Sequential Tests of Educational Progress, Series III	1
	Stanford-Binet	1
	Talent Assessment Checklist	1
	Texas Educational Assessment of Minimal Skills	1
	Watson-Glaser Critical Thinking Appraisal	2
	Wechsler Intelligence Scale for Children-Revised	1

(*continued*)

Figure 1. Standardized instruments used to assess program evaluation questions.

In Figure 2 we present the instruments which were used without specific reference to an evaluation question.

Once the frequency of use was determined for each standardized instrument in each evaluation outcome category, we evaluated each instrument for valid use of the instrument to determine the outcome of an evaluation question or goal and the reliability of that instrument given sufficient evidence of validity. A rating scale and a procedure were developed to evaluate the instruments and evaluation designs using a scale entitled *Scale for Evaluation of Program Evaluation Instruments* (*SEPEI*) (see Appendix B).

Analysis of Evaluation Instruments Related to Evaluation Questions or Outcomes

Local instruments have been catalogued in the EVALNOST database. These locally developed, non-standardized instruments include: assessments of student outcomes such as attitudes toward school and program, content mastery, creativity, independence/responsibility, research skills, risk-taking, self-concept, self-expression, task persistence, and thinking skills. Other factors assessed include: awareness, availability, community/parent involvement, cost effectiveness, counseling, curriculum, enrollment, evaluation, facilities, funding, identification, impact of program on schools, in-service instruction, learning environment, management, materials, non-participant perceptions, participant perceptions, personnel qualifications, planning, program design, program guidelines, program implementation, progress on recommendations, resources, satisfaction, staffing, student/peer interactions, student needs, support, teaching concerns, time, training, and underachievement.

Animal Crackers
Career Decision Making Skills
California Achievement Test
Children's Task Persistence
Kaufman Assessment Battery for Children
Kit of Factor Referenced Cognitive Tests
Piers Harris Children's Self Concept Scale
Preliminary Scholastic Aptitude Test
Role Category Test
Ross Test of Higher Cognitive Processes
Scholastic Aptitude Test
Self-Concept and Motivation Inventory
SRA Achievement Test
TAAS Criterion-Referenced Test (Texas criterion-referenced assessment)
Thinking Creatively in Action and Movement
Torrance Tests of Creative Thinking
Williams Test of Divergent Thinking

Figure 2. List of standardized instruments used but unrelated to a specific evaluation question.

The standardized instruments identified in the evaluation reports are located in the EVALPUB database and number 103. A listing of these instruments, according to evaluation outcome use and frequency of use by outcome, is included in Figure 1.

Of the evaluation reports, 66% (83/126) did not use standardized instruments to measure the outcomes of program evaluation questions. These districts relied on locally developed questionnaires or surveys, interviews, document review, or other qualitative methods to provide the evaluation information. Out of the remaining evaluations that did report using standardized instruments, 28% (36) of the school districts actually used the instruments to assess specific program evaluation questions.

The outcomes evaluated most often using standardized instruments were achievement (12 instruments used by 19 districts), creativity (7 instruments used by 9 districts), identification (16 instruments used by 20 districts), and thinking skills (10 instruments used by 20 districts). The *Comprehensive Test of Basic Skills* was used most often (5 districts) to evaluate achievement outcomes of programs for gifted learners. The *California Achievement Test* was used by 3 districts to evaluate the identification outcomes of programs for gifted learners. The *Ross Test of Higher Cognitive Processes* was used most often (8 districts) to evaluate students' thinking skills after receiving services provided by the gifted program. For creativity, no single instrument was used by more than two districts and most instruments listed were used by only one district.

Some school districts (14) assessed affective outcomes as a result of their program for gifted learners. They were interested in determining attitudes towards others, autonomy/responsibility, locus of control, perceptions, and self-concept. Eight of the 14 districts assessed self-concept using 6 different instruments with 3 districts using the *ME Scale*.

Identifying specific outcomes related to a gifted program proved to be difficult for many districts. For example, several districts (16) identified very broad outcomes that were classified as general academic (5 instruments), general program (7 instruments), or general student growth (4 instruments). The domains tested by these instruments were also very broad and included achievement, aptitude, higher level or critical thinking, school attitudes, basic skills, talent, and creativity. Only two districts identified specific skills as evaluation outcomes (research and study skills). Only one district identified aptitude as a measurable outcome of a program for gifted students.

Of the districts, 6% did not report clearly identifiable program evaluation questions or goals despite the fact that an evaluation was conducted and standardized instruments were used as an assessment instrument in the evaluation. At least 17 instruments were administered without an identifiable evaluation question to define the purpose of the assessment.

Assessing the Psychometric Properties of Published Instruments

The second line of investigation focused on reviewing published instruments which were either cited in journal articles reviewed or included in evaluation reports submitted by school districts, or found in ERIC documents. This phase was subdivided into two parts.

Initially, the staff gathered all available data from the printed literature and from the survey responses on the reliability, validity, examinee appropriateness, norms, usability, teaching feedback, and ethical propriety of the instruments.

These technical data were used to rate each published instrument using a model rating scale developed by project staff, but based on earlier work done by the Evaluation Technologies Program of the Center for the Study of Education and the Humanizing Learning Program of Research for Better Schools, Inc. in their series of test evaluations (Hoepfner et al., 1972; Hoepfner, Strickland, Jansen, & Patalino, 1970). The existing rating scale was modified to reflect the specific uses to which these instruments have been put—addressing a specific evaluation concern or question. The measurement standards of the *Standards for Evaluations of Educational Programs, Projects and Materials* (Joint Committee on Standards for Educational Evaluation, 1981), the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association and National Council on Measurement in Education, 1985), and *Guidelines for Test Use* (Brown, 1980) were used in developing the final tool for assessing the instruments: *Scale for the Evaluation of Program Evaluation Instruments* (*SEPEI*) (see Appendix B). The technical manual for the *Scale for the Evaluation of Program Evaluation Instruments* (*SEPEI*) is found in Appendix C, and inter-rater agreement percentages are included in Appendix D.

Validity and Reliability of *SEPEI*

Content Validity

The initial draft of the *SEPEI* was reviewed by two faculty with expertise in evaluation, two experts in measurement, and two experts in gifted education (all from the University of Virginia). Modifications in criteria and rating scales were made based on their recommendations.

Reliability

In order to assess inter-rater reliability for the *Scale for the Evaluation of Program Evaluation Instruments* (*SEPEI*), four graduate students at The National Research Center on the Gifted and Talented (NRC/GT) at the University of Virginia were asked to rate the *Cornell Test of Critical Thinking*. Following an analysis of the results of this rating and further revision of the rating scales for the items with greatest discrepancy, the same four students rated the *Ross Test of Higher Cognitive Processes*. These tests were selected for evaluation because of the types of goals stated in programs for the gifted—higher level thinking skills and critical thinking. Descriptive statistics are calculated only for instrument items that required a rating of excellent, good, fair, poor, or not applicable. The Kendall Coefficient of Concordance on the *Ross* was computed as significant ($p<.001$).

As noted in Appendix D, (*SEPEI* Inter-rater Item Descriptive Statistics for the *Cornell Test of Critical Thinking* and the *Ross Test of Higher Cognitive Processes*), the standard deviation for items ranged from 0.0 to 1.41 in the ratings for the Cornell. On items that had a rating from all four students, the standard deviation ranged from 0.5 to 1.41. The standard deviation for items ranged from 0.0 to 1.73 on the ratings for the Ross. These data and the significant Kendall were sufficient evidence of inter-rater agreement to give us confidence in the reliability of evaluation instruments using the *SEPEI*.

For each published instrument listed in the Repository, we identified school districts named as the focus in the use of that instrument. Each instrument was reviewed with that question as a focus of the review. Hence, any particular instrument might be rated once, twice, or several times. A total of 78 tests have been reviewed.

Articles, Test Reviews, and Locally Developed Instruments

As noted above, local instruments were also classified according to evaluation questions. Although many instruments were provided to the NRC/GT, none provided any information on the reliability and validity of the instrument. Our guidelines for rating instruments were based on the judgement that instruments lacking evidence of reliability and validity could not be recommended, and hence, would not be reviewed further.

Importance of This Repository

Appropriate program development and modification are based on the collection of valid and useful data on the functioning of a program. Administrators of programs for the gifted have lacked access to instruments which have been validated or even demonstrated to be reliable for measuring most components of their programs. The collection of instruments in a central repository and an evaluation of these instruments by individuals with expertise in evaluation, psychometrics, and gifted education is long overdue in the field of gifted education. Many districts have struggled with the search for such instruments; some have made initial development efforts; some have collected some data on the effectiveness of instruments. The National Repository information provides more general access to a wider range of information by school district personnel.

The purpose of the databases is to allow practitioners to summon information on instruments other school districts are using to evaluate gifted programs and to access information on the qualities of particular instruments. A sample of response to a request for a search is presented in Appendix E.

CHAPTER 2: Literature Review

Setting the Stage for Focusing on Evaluation

Numerous reasons exist for evaluations, among them: improving effectiveness of programs and program personnel, reducing uncertainties, assisting with decision making and goal setting, seeking justification for decisions, meeting legal requirements, fostering public relations, enhancing the professional stature of the evaluator or program administrator, boosting staff morale, mustering program support, and changing policy, law or procedure (Alkin, 1980; Bissell, 1979; Mathis, 1980; Ostrander, Goldstein, & Hull, 1978; Raizen & Rossi, 1981). Nonetheless, the literature of education is replete with examples of evaluation findings that never resulted in program enhancement, improvement, or development. Disregard for findings of educational evaluation is costly in effort, monies, and in human terms when potential program improvements are stillborn (Datta, 1979; King, Thompson, & Pechman, 1981).

Because of a general lack of public understanding of and support for programs for the gifted, and keen competition for scarce resources, the survival of programs for gifted learners may depend on carefully planned evaluations which yield useful information that can be translated into documentation of effectiveness and action to improve programs by educational decision makers (Dettmer, 1985; Renzulli, 1984). Gallagher (1988) has included program evaluation among the priorities he identifies as crucial for the continued improvement of gifted education. Gallagher states . . . "We risk losing fair documentation of the genuine contribution that such programs [gifted] make if we cannot come forth with a general strategy of how to design appropriate evaluation programs and assessment procedures for these special groups" (p. 112). However, this is problematic as evaluation information is scant for the field of gifted education, even though the call for improved evaluation of programs for the gifted is certainly not new. Gallagher, Weiss, Oglesby, and Thomas (1983) indicate that as early as 1960, when accountability and evaluation were identified as important components of educational programs, the call for evaluation of gifted programs was included. Despite identification of issues and ways of addressing those issues (Callahan, 1983; Renzulli, 1975), the continued call for revisions in the process (Callahan, 1984; Callahan & Caldwell, 1986), and the demand for such undertakings, a national survey by Gallagher et al. (1983) yielded only scanty reports of program evaluation efforts.

In Chapter 1 we discussed the compilation and rating of instruments used in the evaluation of programs for the gifted and the creation of a National Database for accessing information about the use of those instruments. A summary of this investigation is also available in the *Journal for the Education of the Gifted* (Callahan & Caldwell, 1993). In this chapter we will present the results of a synthesis of literature on evaluation utilization and the evaluation of gifted programs.

Evaluation Utilization Rationale

Talk about educational evaluation is plentiful, resources invested in it abundant; yet the literature of education is saturated with examples of non-use of its findings. Issues which surround utilization of results of educational evaluation are numerous and complex. There are questions of definition and philosophy, questions of process and method, and general questions of utility. It is important to understand those central questions as they pertain to educational evaluation in general and as they apply to the field of gifted education in particular so that both the quality and utility of evaluation can be informed and enhanced. These evaluation studies undertaken by The National Research Center on the Gifted and Talented at the University of Virginia examined issues related to use of evaluations and evaluation findings in general, and how those issues have been translated into evaluation of programs for gifted learners.

The literature review was undertaken as background for the focus of the evaluation component of the overall project ascertaining the effectiveness of various evaluation models and strategies. As part of that effort we identified the characteristics of effective evaluations that provided data that proved useful for bringing about change in gifted programs. While the primary research activities obviously were to document characteristics of designs yielding evidence perceived as useful by evaluation audiences in program development and modification, a related purpose was to identify and/or develop guidelines for evaluations that would provide the most accurate, timely, and useful information for policy development and program improvement. The results of this research should provide guidance for schools and evaluators in implementing quality evaluation—defined by us in the project as evaluation that is perceived as useful and actually used for program development purposes.

To achieve our goals, we first reviewed the extant literature dealing with evaluation utilization in general and the literature on evaluation as applied to programs for the gifted and talented. This literature review supported the two distinct but interrelated studies that were conducted. The literature review has been shared with the public in Tomlinson, Bland, and Moon (1993). In the first study, Hunsaker and Callahan (1993) examined the existing trends in the evaluation of programs for gifted and talented students. Based on this information, Tomlinson, Bland, Moon, and Callahan (1994) examined the ways school systems utilized information gathered from evaluations of gifted programs. See Appendix F: A Planning Guide for Evaluating Programs for Gifted Learners and Appendix G: Guidelines for Conducting Useful Evaluations of Programs for Gifted Learners.

Data Gathering

A search of educational databases was conducted to find reports available in the professional literature. Database searches included VIRGO (the computerized card catalogue system of the University of Virginia), ERIC, PsycLIT (the computerized version of *Psychological Index*), and *Dissertation Abstracts International.* Search terms included evaluation, design, implementation, utilization, and gifted. These terms were used singly and in combination as appropriate. Each search yielded a list of potential references that was reviewed; those identified as promising resources were located and placed into the appropriate database. Fifty nine general evaluation articles, 38 evaluation utilization articles, and 15 evaluation design articles were identified and abstracted. Fourteen articles dealing with evaluation of gifted programs were also identified.

Theoretical arguments and empirical findings were synthesized to provide an overview of current knowledge about practices which influence the degree to which data collected as part of the evaluation process are used and used appropriately in decision-making relative to program improvement.

Definitions

Definitions of "utilization of evaluation results" span an impressive range from narrow and restrictive (a single intended user of results making a specific decision immediately upon receipt of findings, and basing the decisions heavily upon those findings), to broad and vague (anyone using anything from an evaluation report). Evaluators also realize that "use" can range from concrete action such as making decisions about a program, to a more abstract or conceptual response such as altering one's thinking about a program (Alkin, 1980; Alkin, Dailak, & White, 1979).

Patton (1988) links evaluation and its utility when he describes evaluation practice as a "systematic collection of information about the activities, characteristics, and outcomes of programs, personnel, and products for use by specific people to reduce uncertainties, improve effectiveness, and make decisions with regard to what those programs, personnel or products are doing and affecting" (p. 301).

Purposes for Evaluating

Evaluation is used at local, state, and national levels by people in widely varying roles from program staff to school boards and from parents to funding agencies. Given the wide range of users, it is not surprising that the literature suggests evaluations are conducted for many reasons, among them to: seek justification for decisions, meet legal requirements, foster public relations, enhance professional prestige for the evaluator or administrator, encourage continuation of successful program components, inform decision-makers and funding agencies, share information with a varied spectrum of audiences, boost staff morale, build program support, modify laws or regulations, and influence curricular choices or strategies (Alkin, 1980; Bissell, 1979; Mathis, 1980; Raizen & Rossi, 1981). David (1981) noted that findings from Title I evaluations were used primarily "to meet legal requirements, provide feedback, and provide gross indicators of program effectiveness" (p. 31). Patton, Grimes, Guthrie, and Brennan (1977) found that educational evaluation "is used by decision-makers, but not in the clear-cut and organization shaking ways that social scientists sometimes believe research should be used" (p. 144). These reasons for evaluations typically fall into two categories: program improvement (e.g., seeking information for program improvement) and program protection (e.g., meeting legal requirements).

Mathis (1980) notes there may be more than one reason which spurs the initiation of the evaluation process. Because the explicit and implicit impetuses may be mutually exclusive, there is a need to ensure that purposes of evaluation are always made explicit and that conflicts in purpose are resolved prior to implementation of the evaluation process.

Evaluation Design

The subject of evaluation design is complex, and a topic explored fully in book length discussions and textbooks. Because design will shape information and the shape of information will impact utility, it is important to note here that the issue of selecting an appropriate evaluation design is perhaps more ardently debated than any other subject related to evaluation. At one end of a continuum are those who argue that the use of quantitative experimental design for evaluation of social programs (Fairweather, 1981) is imperative. Others contend that qualitative methods are uniquely suited to the complex and multi-faceted nature of the educational endeavor (Guba, 1978; Patton, 1989, 1990). In between, lies the belief that the strongest evaluation designs will consist of a combination of experimental and non-experimental methods. Smith (1980) suggests that experimental design is desirable when dealing with causal questions, looking at a narrow range of program variables, examining an established program, and when contextual factors are unimportant. Non-experimental methods are preferred when conducting an exploratory set, dealing with a broad range of questions, or evaluating an emergent program. Thus it becomes important to select an evaluation design according to the context to be evaluated rather than out of dogged adherence to either a positivist or phenomenological paradigm.

Factors Affecting the Use of Evaluation Findings

Other factors beyond evaluation design that are also assumed to affect utilization of evaluation information may be divided into (1) factors pertinent to the context of the evaluation but *out of* the evaluator's control, and (2) those factors, at least to some measure, *within* the evaluator's control.

Evaluation Context Factors

Factors associated with evaluation context can be further divided into economic concerns and political factors. Of the two, economic concerns are paramount, for without funds to enable implementation of recommendations, utilization is impossible (Marshall, 1984; Patton, 1988). Even resources and funding, however, hinge on the political environment. If there is a lack of prior public commitment to a program or to suggested changes, utilization of findings can be controlled politically (Brown, Newman, & Rivers, 1980; Marshall, 1984; Patton, 1988).

Mathis (1980) believes that educational evaluation is nearly always political in nature because educational programs are generally political creations. Biases in the use of evaluation results may evolve from such politically based variables as the individual or group who generated the evaluation, the reasons why the particular program was selected for evaluation, the selection of particular program objectives to consider for the evaluation, the selection of a person or persons to conduct the evaluation, and level of support for the evaluation activities. While we value rational decision-making in the abstract, says Mathis, the use of evaluation data is often selective and serves to further political ends. In addition, he cautions that evaluation results are seldom as clear cut as policy makers would like and indicates that the subtleties, cautions, and caveats of evaluation findings are often lost as evaluation findings become politicized.

Evaluator Control Factors

The Joint Committee on Standards for Educational Evaluation (1981) established guidelines for evaluations. These guidelines speak to evaluation elements which, unlike economics and political climate, are subject to evaluator control. They include utility, feasibility, propriety, and accuracy. The critical importance of the utility standards, stems from the assumption that evaluation should not be conducted if utilization is not going to occur. Appendix H provides a summary of articles relating to utilization classified according to the Joint Committee utility standards of audience identification, evaluator credibility, information scope and sequence, valuational interpretation, report clarity, report dissemination, report timeliness, and evaluation impact. Braskamp, Brown, and Newman (1981) suggest grouping these variables into the larger domains of message source, message content, and characteristics of the receiver.

Message Source

The first of these categories, message source, includes evaluator credibility and valuational interpretation. In relation to message source, Braskamp et al. (1981) reported that readers were less likely to agree with reports they thought were written by female evaluators. In addition, readers felt reports were more objective if they were written by a "researcher" rather than an "evaluator" or a "content specialist."

Message Content

The message content category encompasses information scope and sequence as well as timeliness and clarity of the report. The manner in which evaluation results are presented to potential users obviously affects the audience's comprehension of the message. And understanding will obviously affect the extent and the appropriateness of use of the findings (Gold, 1983; Patton, 1988). Several findings in this regard offer direction to evaluators seeking to ensure use of evaluation findings. First, the use of research jargon is rarely appropriate in communicating with decision-makers. Further, readers state that those reports which combined use of jargon with statistical data are the most difficult to read (Bickel & Cooley, 1981; King et al., 1981). Similarly, the use of statistical data in combination with other features such as technical language, excessive report length, and inclusion of negative results may have a negative impact on audience reaction (Brown, Newman, & Rivers, 1980). Clients, specifically administrators, prefer qualitative rather than quantitative information (Alkin & Stecher, 1981).

Characteristics of the Receiver

Characteristics of the receiver include audience identification, report dissemination, and evaluation impact. In an evaluation study that employed qualitative methodology, D'Amico and Dawson (1985) used a research approach to assess particular recommendations relative to evaluation utilization. They found quick turnaround, use of client-centered feedback strategies, and involvement of clients directly with data collection and analyses increased utilization of findings. Bickel and Cooley (1981) concluded that a clearly identified client and frank dialogue with the client throughout the evaluation process (evaluation design and implementation phases) increased chances that the evaluation findings would be used. "Those individuals with a high perceived need for evaluation were generally more satisfied with the information they had available than those with a low perceived need" (Kennedy, Apling, & Neumann, 1980, p. 11-12).

Communication in General

In addition to the three categories suggested by Braskamp, Brown, and Newman (1981), effective communication methods seem to have a positive impact on utilization. Effective communication with users of the evaluation results provides for education about the evaluation, its recommendations, and its utility. Such communication can foster rigorous thinking about the evaluation, whether or not immediate implementation of recommendations occurs, and it thus, can lead to long-term benefits. Stake (1975) and Gold (1983) both call for more user involvement in the evaluation process. Stake's "responsive evaluation" approach calls for evaluators to consult users and incorporate their interests and concerns into the evaluation design if possible. Gold's "stakeholder" approach encourages evaluators to adhere to user preferences for both the type of information the audience desires and the forms in which they wish to receive the information.

Use of Multiple Data Collection and Reporting Strategies

A final factor influencing evaluation utilization involves data collection and reporting. Turner, Hartman, Nielsen, and Lombana (1988) surveyed decision-makers in four evaluation studies which they (Turner, Hartman, Nielsen, & Lombana, 1988) had conducted and in which they had used multiple data gathering methods. The purpose of the study was to determine (1) the degree of use of recommendations and (2) factors affecting the use of recommendations. These researchers found that the two chief categories impacting evaluation utilization were timeliness of reporting and substance of the report. They also found that use of multiple data gathering methods was among the top five reasons given for utilization and was directly related to other utilization factors such as the evaluator's willingness to involve users, rapport with users, evaluator credibility, user's commitment to use of results, substance of evaluation information, and evaluation reporting. The authors hypothesize that using a wide variety of data gathering methods facilitates close communication between evaluator and program participants, and builds trust in the evaluation findings. Furthermore, it affects the credibility of the evaluators as it allows them to display skills in varied ways, allows visibility with a more diverse set of potential audiences, increases evaluator understanding of a project, enables the evaluators to be more fluent in answering questions about the study, encourages a mix of data collection sources and reporting methods, and perhaps most importantly, lends credibility through triangulation of results.

A review of the general literature of educational evaluation thus indicates that utilization of results may be affected by factors of design, economic and political contexts, and degree of adherence to utility standards. Some of the factors which have an impact on utilization of findings are, at least to a degree, under evaluator control. Some are not. Evaluators may positively influence those factors which yield the most promise for improved evaluation designs and use of evaluation findings and recommendations.

Utilization of Evaluation Results in Gifted and Talented Programs

The literature of evaluation as applied to gifted education is scant. The literature on empirical evaluation utilization as it relates to gifted education is virtually non-existent. However, the literature does suggest issues and concerns which relate to utilization and which should inform practice of evaluators of gifted programs.

Special Challenges in Evaluating Gifted Programs

Programs for gifted learners have several characteristics which confound evaluation and subsequently constrain the use of evaluation findings by virtue of the fact that the evaluation design itself may produce weak findings. The very goals of programs for the gifted render the evaluation process difficult. They most often can be characterized as holistic, complex, long-term, product-oriented, and individualized; hence, they are difficult to measure in traditional ways using traditional assessment tools (Callahan, 1983; Ganapole, 1982; Gilberg, 1983; Renzulli, 1984). For example, it is difficult to develop adequate evaluation constructs for programs which seek to develop creativity in students over an extended period of time. It is difficult to quantify the progress of a specific lone fifth grader working with materials relating to preservation techniques in archeological digs.

The obvious shortcomings of standardized measures as tools of evaluating programs for the gifted have led to an inordinate reliance on attitude surveys which are easy to construct and administer, and which provide non-threatening information. (See Table A-5 in Appendix A for the data from our analysis of evaluation report for further confirmation of these assertions.) As a result, there has been little use of outcome indicators and, sometimes, use of measures which are invalid, unreliable, or just unrelated to program content.

Evaluation designs used in assessing programs for gifted may err by focusing on short-term goals when evaluation of long-term goals would be more appropriate. Callahan

(1983) also points out, in fact, that we are not even certain of the validity of the evaluation questions we ask in such settings. She notes that the problems which surround the evaluation of program goals for gifted learners are complicated by the fact that there exist no agreed upon standards of "good programming," and no common set of standards for student performance against which achievement may be assessed. Behavioral objectives, a strategy used to establish standards of achievement for Title 1 or special education programs, have often been too vague, narrow, or otherwise inappropriate for assessing the progress of gifted learners when adapted to programs for the gifted (Callahan, 1983).

The problems presented by the complex and individualized nature of goals of gifted programs are accompanied by measurement problems. Standardized measures have been largely ineffective in evaluating gifted programs for several reasons. Gifted learners are identified at least in part by their scores, which are at the top of standardized tests. Subsequent testing which might ordinarily be employed to examine academic growth will most likely, for gifted learners, result in lowered scores due to regression to the mean, or the tendency of high (or low) scores on a test to move toward the middle of the score range when the test is readministered to high (or low) scorers. Low ceilings on such tests are accompanied by deceleration of gains for older students in general, and thus, present a dual problem in assessing older, gifted learners. Further, few if any standardized tests are constructed to measure the advanced or complex sorts of learning encountered by gifted learners in settings with appropriately differentiated curricula. There are no norms for gifted learners per se on most standardized measures. Finally, reliability of standardized test scores typically decreases with the increased homogeneity of the group being measured, and gifted learners are a relatively homogeneous group (Callahan, 1983; Gilberg, 1983; Renzulli, 1984).

While problems of broad and long range goals may be combated by using tests and scales with sufficient range to show growth over time and ceiling effects may be offset by using out-of-level tests or tests normed on older populations (Beggs, Mouw, & Barton, 1989), the complexity and abstractness of content is not addressed in currently available standardized instruments, so that validity of assessment remains a fundamental challenge.

The evaluation of gifted programs suffers from lack of focus resulting from poorly articulated program goals measured by instruments that are ill-suited for the purpose. This may result in altering instruction solely for the purpose of attaining higher test scores (teaching to the test) or providing amusement rather than challenge in order to raise ratings on "attitude toward this program" assessments.

Recommended Evaluation Designs

As is the case in the general literature of evaluation, the literature of evaluation of gifted programs lacks concurrence with regard to the desirability of quantitative vs. qualitative methods, or the importance of using experimental design. Difficult issues arise from implicit assumptions in experimental design. First, as in medical experimentation, educators question the assignment of subjects to control groups. While we most often do not know the actual impact of a program, many judge it inappropriate to exclude students from an intervention *believed* to be positive. Second, there is an additional expense of implementing alternative forms of an intervention if a comparison design is used. Practical constraints may necessitate inclusion of all students in the intervention groups. Selection procedures, knowledge of treatment, and the John Henry effect (members of a control group work especially hard in order to compete with an experimental group) (Callahan, 1983; Payne & Brown, 1982) also hamper efforts to establish control groups.

Several modifications to traditional experimental design have been proposed. In lieu of traditional "control groups," use of "contrast groups" (Payne & Brown, 1982) or "comparison groups" (Beggs et al., 1989) are proposed. These are existing groups or to-be-generated data sets against which the results of a particular intervention may be contrasted.

Payne and Brown (1982) and Carter (1986) suggest two alternatives to randomization in experimental design: (1) use of an Aggregate Rank Similarity contrast group derived from a judicious matching of schools, classes or school systems, and (2) use of retrospective pretesting in which a group serves as its own control through a backwards look at how they have changed as a result of treatment. In this process, group members answer questions after treatment about their skills and/or knowledge as they would have answered them prior to treatment and again as they would answer them following treatment.

As an additional alternative to randomization, Callahan (1983) suggests using a time-series design in which several groups receive the intervention in question, but at various times in a year, thus allowing the groups to serve as controls for one another. Further, she proposes use of students as their own controls in instances when students may rotate in and out of programs for a variety of personal or programmatic reasons. Carter (1986) suggests providing the same intervention to classrooms of non-gifted learners as well as classrooms of gifted learners to determine the breadth and depth of achievement and rate of learning of the two groups in order to better understand the effects of differentiated education.

While reiterating the need for outcome-based evaluation in programs for the gifted, Carter and Hamilton (1985) propose that quantitative designs are appropriate for outcome-based evaluations, while qualitative designs are appropriate for process-based evaluations (i.e., examination of documents related to a program via content analysis). Qualitative evaluation methods are more broadly commended and viewed as especially well suited to evaluation of gifted programs by Lundsteen (1987) because they assist in understanding the processes in which gifted learners and their teachers are involved, help in establishing meaningful hypotheses for further study, and avoid the error of oversimplification of complex settings and procedures. Janesick (1989) likewise sees utility in qualitative methods because they allow evaluators to look at multiple realities, and they are useful in establishing a process which is change oriented and educative for a variety of stakeholders. She suggests gathering three types of data: baseline data (about the research setting, participants, demography, etc.), process data (which describe what happens during the course of curricular innovations being studied), and values data (which yield information about the values of various stakeholders and which of those values the program in question supports and neglects).

To facilitate use of evaluation findings, it is imperative that evaluators of gifted programs select evaluation designs appropriate to the evaluation focus and context so that findings will be both useful and meaningful. Designs which yield findings that appear inconsequential will be unlikely to merit serious attention from policy makers who have the power to translate findings into action.

Utility Standards and Evaluation of Gifted and Talented Programs

In addition to guidance regarding evaluation method and design which will influence utility of findings, writers in the field of program evaluation in gifted education provide other suggestions for evaluation of gifted programs which roughly parallel some of the

Joint Committee utility standards grouped according to message source, message content and characteristics of the receiver.

Recommendations Related to Message Source

In evaluating gifted programs, there is a need to ensure that both staff and evaluators are trained to carry out the evaluation and analyze the results of evaluations of programs for gifted learners (Gilberg, 1983). There is also a need to prepare and describe scoring rules prior to administration of tests (Ganapole, 1982).

Recommendations Related to Message Content

It is essential to understand that an evaluation mirrors the presence or absence of appropriate program structures and goals, and that evaluations cannot succeed if these elements are lacking or inadequate (Dettmer, 1985) or if the program structures or goals are not fully and appropriately addressed in the evaluation process. Evaluation will also be enhanced and the chances of findings being used will increase if concerns of both internal and external audiences of programs for the gifted are clearly addressed.

According to Callahan (1986), if questions which are relevant, useful, and important form the foundation of the evaluation, the evaluation will be enhanced. Relevant questions are those which clearly address the function, components, goals, activities, and structure of the program. Further, evaluation questions are not research questions, hence they are relevant to a particular program, not the field in general. Evaluation seeks specificity, not generalizability. Useful questions are those which provide data that an audience can actually use in decision-making. Important questions are those that will yield data helpful in making decisions that can have a significant impact on programs and participants. She suggests asking evaluation questions relating to these areas of the program which: are of central importance to program effectiveness, present potential problems, concern availability and adequacy of resources, address areas that might result in undesirable change, reflect conflict with general institutional values, may cause individual loss of power, present economic threat, deal with potential for inconsistency between suggested action and actual action, may uncover lack of understanding of goals, and reflect the personal bias of significant audiences. Finally, it is important to employ varied data collection modes in response to the needs of varied constituencies of gifted programs (Gilberg, 1983; Janesick, 1989; Rimm, 1982).

Recommendations Related to Receiver Characteristics and Audience Identification

Use of evaluation findings will be encouraged if decision-makers at various levels are identified. Of equal importance is an understanding of the actions over which they have control (Callahan, 1986; Dettmer, 1985; Gilberg, 1983; Renzulli, 1984; Rimm, 1982). Of course, this knowledge will only be of value if the evaluator ensures that information relevant to particular decisions reaches the decision-maker who has the power over the final adjudication of that issue.

Gilberg (1983) encourages evaluators to find out what courses of action will result from data supplied, and to make recommendations with an eye toward program improvement. Dettmer (1985) suggests that maximum impact will result if self-studies are conducted by local advisory councils based on evaluation data, recommendations are made as a result of the self-study, reports are prepared for each stakeholder group, and specific actions for carrying out the recommendations are discussed and procedures for implementation are initiated.

Applied Evaluation in Gifted and Talented Programs

A search of professional journals yielded only one example of an *evaluation utilization* study specifically applied to programs for gifted learners. Turner et al., (1988) studied evaluation utilization following a three-year evaluation process in a program for academically gifted learners. Of 32 recommendations made in the evaluation report, 23 (72%) were acted upon. A variety of data gathering methods were used in the evaluation itself including mail-out surveys, telephone and face-to-face interviews, classroom observations, town meetings, paper and pencil tests, record reviews of science fair entries and class rosters, and staff development offerings. The authors conclude that evaluation utility and comprehensiveness were a direct result of the use of multiple data gathering methods.

It is important that educators of the gifted examine evaluation utilization in the field of gifted education according to accepted utility standards as related to message source, message content, and receiver characteristics to develop an understanding of utilization factors which are both within and beyond evaluator control. Such a systematic study would undoubtedly clarify factors and constellations of factors which constitute effective evaluation designs for these unique programs. Furthermore, it would ensure that worthwhile evaluations are conducted, and increase the likelihood that meaningful actions follow.

CHAPTER 3: Current Practices in the Evaluation of Gifted Programs

The literature of gifted education is nearly mute on evaluation utility. There is a clear need in the field of gifted education to address the difficult issues of evaluation which directly influence positive and appropriate use of evaluation findings if programs for the gifted are to achieve educational rigor and continued development. In times of limited resources for educational programs, the survival of services appropriate for gifted learners may depend on carefully planned and comprehensive evaluations that document all aspects and outcomes of services, and yield useful information for decision-makers to improve program effectiveness and improve the cost/benefits of programs (Dettmer, 1985; Renzulli, 1984). While the utilization of evaluation findings for program improvement is important and a shared desired outcome of all program evaluations, the utilization of evaluation findings in programs for gifted students serves another important function. Because programs for gifted learners do not usually enjoy popular support for a variety of reasons, it is all the more essential that educators be able to demonstrate solid student growth for participants. To the degree that a positive ripple effect for the entire school is documented, the program has potential to gain increased support from the general community. It behooves those offering services to gifted students to use evaluation data to demonstrate that the programs are resulting in change (rather than wasting limited resources) and that the results of evaluation are used to enhance and make the program more efficient. If the program is not resulting in desired outcomes, that information is also vital in our considerations of how to best meet the needs of highly able students.

The literature review we conducted identified the relative paucity of information dealing with gifted program evaluation. It did not specifically describe the current practices school system personnel use to evaluate gifted programs or how school districts utilize that information for program improvement. In an effort to determine current gifted program evaluation practices, we conducted a review of 70 evaluation reports collected by The National Research Center on the Gifted and Talented at the University of Virginia from public and private school and professional sources.

Methodology

Data Gathering

Gifted program evaluation reports were collected from three sources: a search of educational databases; an appeal through professional journals, newsletters, and conferences for submission of such reports; and direct mail requests to state-level gifted coordinators, school districts which had indicated an interest in collaborating with the research, and approximately 5,000 other individual school districts.

The search of the educational databases was conducted to find reports available in the professional literature. Database searches included VIRGO (the computerized card catalogue system of the University of Virginia), ERIC, PsycLIT (the computerized version of *Psychological Index*), and *Dissertation Abstracts International*. Search terms included gifted ratings, scales, tests, measurements, evaluation, and utilization. These terms were used singly and in combination as appropriate.

An appeal for copies of evaluation reports was made through journals, newsletters, and conferences. Searches were conducted to determine the professional organizations, journals, and state associations through which it would be appropriate to make requests for information. Newsletter releases were prepared and mailed to each organization, journal, or association. Individual requests to members of the National Association of Gifted Children and the American Evaluation Association were made through inserts in their annual convention packets.

The final special mailing to state-level gifted coordinators, NRC/GT Collaborative School Districts, and to approximately 5,000 local school districts was conducted in conjunction with the request for information for identification information. The addresses for these districts were obtained from an educational database firm. Where possible, alternative means to the postal system were used to distribute these requests. These were done through state conferences (Florida-75 letters, Iowa-540 letters, Virginia-175 letters), state associations (Texas-1068 letters), and state gifted coordinators (Arizona-96 letters, Colorado unknown).

Data Analysis

The 70 evaluation reports we collected were coded by NRC/GT staff trained in gifted program evaluation on 10 variables: evaluation type, evaluation model, evaluator type, data-gathering methodology, data analysis technique, data sources, intended audiences, reporting format, evaluation concerns, and utility information. Predetermined categories within each variable were based on a review of evaluation and gifted education literature. The definition of evaluation terms essentially followed those given in Worthen and Sanders (1987).

Frequencies of the categories within each variable were computed. Further, in order to determine the independence of the variables, chi square analyses were conducted on data from all pairs of combinations for 9 of the 10 variables. The variable "evaluation concerns" was not included due to difficulties with inflated N (that is, no meaningful categorization of the data was possible while still maintaining independent observations for each evaluation report). More sophisticated analyses were not conducted due to the non-parametric nature of the data.

Limitations

The data reported here are indicative of trends in gifted program evaluation only. Due to non-random sampling, generalizations should be made cautiously.

Further, chi square analyses should be interpreted cautiously. With the large number (36) of analyses conducted, it is possible that a single analysis could exceed the critical chi square value at the .05 level due to chance alone.

Results

The result of this study was a description of current general trends in evaluation of programs for gifted learners. The tables in Appendix A provide specific details for the frequency and chi-square analyses.

A Typical Evaluation

Based on the frequency analysis we can characterize a typical evaluation of a gifted program as a summative evaluation focusing on multiple concerns raised by program or school central administrators and conducted internally rather than by an external evaluator.

Nearly all data are collected by questionnaires with relatively infrequent use of tests, document analysis, observations, or focus group meetings. Data are reported most frequently using descriptive statistics alone. The use of multiple sources of evaluation data prevails and students, parents, and teachers are most often the sources of data with governing bodies and counselors rarely involved!

The report is written for administrators and focuses on concerns about curriculum, identification, program organization, and general impressions of the programs. The report will provide a general narrative and includes tables of frequency distribution and statistical analysis, if inferential statistics are used. Rarely will there be an executive summary. Surprisingly, it will not include recommendations in more than half the cases. Rarely are there any other provisions for abetting the implementation of recommendations (e.g., timelines, task definitions and assignments, or policy and goals formulation).

Specific Details of the Frequency Analysis

Evaluation Types

Three basic evaluation types were employed in the reports analyzed; summative to determine program worth, formative to improve the program, and needs assessment to determine the need for a program. The most frequently reported type of evaluation was summative evaluation (55.7%). Formative evaluation was included in 35.75% of the cases. Two districts reported combining elements of summative and formative evaluation. Other types of evaluation were the focus of 7.1% of the evaluations. (See Appendix A for Tables.)

Evaluation Models

Four categories of evaluation models were employed. Management-centered evaluation (the evaluation concerns of program administrators are addressed) was used in 57.1% of the reports. Objectives-centered evaluation (the goals and objectives of the program are addressed) typified 28.6% of the reports. Product-centered evaluation (focusing on the value of a specific gifted program model for possible adoption or transfer) was the focus in 14.3% of the cases. Participant-centered evaluation (focusing on concerns and perception of all program participants) was used in only 5.7% of cases. Three reports (4.3%) combined these models.

Evaluator Types

Most of the evaluations were conducted by an internal evaluator (58.6%). External evaluators were responsible for 42.9% of the reports. In only one case did the report combine the efforts of external and internal evaluators.

Data Gathering Methods

Both quantitative and qualitative methods of gathering data were reported by districts studied. Only the questionnaire was used in a majority (77.1%) of the evaluations. Other frequently used data collection strategies included testing (37.1%), document analyses (32.9%), observation (31.4%), and interviewing (30%). Meetings (11.4%) and other methods (e.g., clinical analysis, product ratings) (7.1%) were less frequently used. Most reports (61.4%) used a combination of methods for gathering data.

Data Analysis Techniques

While both quantitative and qualitative analysis techniques were reported, descriptive statistics definitely dominated the analyses of data in the reports (62.9% of cases). Data were analyzed with inferential statistics 24.3% of the time. Content analysis was the predominant (32.9%) qualitative analysis technique employed. In 11.4% of the reports, data were reviewed in light of professional standards. Other qualitative analyses (e.g., ethnography, impressionistic, narrative) characterized 22.9% of the reports. Multiple methods of analysis were used in 42.9% of the evaluations.

Data Sources

In almost half of the reports (41.4%), students (75.7%), parents (61.4%), and teachers (61.4%) provide most of the data. Input from governing bodies (e.g., school boards) and counselors was sought in only 8.6% and 2.9% of reports respectively. Over 75% of the reports indicated that multiple data sources were tapped.

Intended Audiences

The majority (75.7%) of evaluation reports in this study were written for administrators. The second largest audience (25.7%) was the research community (i.e., other researchers, evaluators). Other intended school audiences included the governing body (15.7%), teachers (8.6%), and counselors (2.9%). Parents were considered an audience in only 4.3% of the reports. About 25% of the reports were written for multiple audiences.

Evaluation Concerns

The evaluations most often addressed multiple concerns. Only 12 of the reports focused on only one concern. The mean number of concerns per report was 4.1. The most frequent areas of evaluation concern were curriculum and instruction (52.9%), identification (44.3%), organization (e.g., models, schedules) (44.3%), and parent/community involvement (42.9%). General impressions of the program was a concern in 42.9% of the reports (e.g., My child is challenged. My child enjoys the program). Measurement of specific program outcomes was characteristic of only 37.1% of the reports. Staff development issues (e.g., teacher selection, training, evaluation) were dealt with in 35.7% of the reports. Student adjustment (e.g., problems, counseling needs) was dealt with in 32.9% of the reports. Less than a third of the reports looked at resources (funding, facilities, materials), underserved populations (minorities, underachievers, learning disabled), program foundations (philosophy, goals/objectives, definition), and program and student evaluation.

Reporting Formats

General reports (65.7%) and data tables (64.3%) were the most frequent reporting formats. Executive summaries characterized only 27.1% of the reports analyzed. Other reporting vehicles, such as oral presentations, memoranda, and journal articles, were evident in 17.1% of the evaluations. Just over half the evaluations used multiple reporting formats.

Utility Practices

Utility practices are those activities of the evaluator designed to increase the likelihood that evaluation information will be useful in generating program policy or improvement. Approximately 43% of the evaluations contained recommendations only. Some reports (27.1%) went beyond recommendations to produce time lines for implementation, task definitions, and policy and goal formulations. In some cases reports indicated that committees were formed for the purpose of implementing recommendations. Of the reports, 30% included no utility information.

Specific Chi Square Analyses

In order to avoid expected frequencies of less than one, a number of categories within each variable were collapsed in the chi square analyses. Of the 36 chi square analyses conducted, 15 had significant results at or above the .05 level. Only the essential differences between observed and expected frequencies of the significant analyses are discussed here. Consult Appendix A for the specific statistical tables associated with these results.

Internal evaluators were more likely to use a management-centered evaluation model than any other model. On the other hand external evaluators, were likely to employ other models, such as product-centered evaluation.

Internal evaluators were more likely to use questionnaires as a data gathering methods than were external evaluators. Also, external evaluators were more likely to use multiple data gathering methods than to use any particular data collection strategy by itself.

In comparing frequencies with the chi square analyses, we note that counselors, governing bodies, teachers, and other sources were included in evaluations only as part of a multiple data source scheme. Most data gathering with multiple methods drew from multiple data sources. Not surprisingly, when students were the only source of information, then data gathering methods such as tests, were more likely to be used. When the sole source of gathering information was parents or administrators for example, they were more likely to complete questionnaires as the only data gathering methodology.

When the intended audience was solely administrative, evaluations tended to be summative, though the formative evaluation was also used (Table A-16, Appendix A). The dominance of summative evaluations was even more striking for research audiences. When the intended audiences were multiple, formative evaluations tended to be favored.

Further, exclusively administrative audiences were more likely to receive the results from management-centered evaluations than from evaluations conducted with other models. On the other hand, evaluations focused on research audiences favored other models, particularly objectives-centered and product-centered evaluations.

When considering intended audience by evaluator type, we note that administrative audiences tended to be associated with internal evaluators more and external evaluators less. In many cases internal evaluators were the gifted program administrators themselves. Also, the research community received information more frequently from external evaluators.

Research audiences were more likely to receive information from reports using quantitative data analysis techniques. When multiple audiences were the target, results that combined quantitative and qualitative analyses dominated.

Multiple reporting formats characterized most evaluations, and were based on the evaluation model used. However, in management-centered evaluations we were more likely to encounter only tables than we were in other models. Objectives-centered evaluations produced fewer evaluations with tables only.

Internal evaluators were more likely to use tables as the sole reporting format, whereas external evaluators were more likely to use other reporting formats (e.g., journal articles, executive summaries). In both cases, however, multiple reporting formats were used most often.

Multiple data-gathering methods yielded multiple reporting formats. The use of tables only was rare when multiple methods were used. The use of only questionnaires for data gathering tended to produce reports with tables used alone and was less likely to result in multiple reporting formats.

Evaluations in which qualitative analyses were used were more likely to result in general reports and other reporting formats, but were less likely to result in multiple reporting formats. Evaluations using quantitative analyses alone or in combination with qualitative analysis were usually reported through multiple formats.

We found that the primary reporting format, regardless of intended audience, was multiple formats. However, administrative audiences tended to get more reports with tables only. Evaluation associated with research audiences were characterized as using multiple format reports, but were also more likely to receive reports using "other formats," such as journal articles.

Summative evaluation was more likely to be associated with reports that lacked basic utility information. Further, a report going beyond recommendations (e.g., giving an action plan, reporting policy development) was less likely to be a summative evaluation. The opposite was true for formative evaluation, from which reports were more likely to give information going beyond recommendations.

Essentially, evaluations employing multiple methods were more likely to yield reports which included utility information. These evaluations yielded a number of reports with recommendations as a minimum, including a number of reports going beyond recommendations. Use of questionnaires as the sole data gathering method tended to result in reports with no utility information.

Finally, making recommendations as a minimum, and often going beyond recommendations, was associated with reporting evaluation results in multiple ways. On the other hand, using tables as the only reporting format was less likely to yield recommendations and more likely to be associated with having no utility practices reported.

Concerns and Promising Practices

Among concerns noted in this phase of the study were an apparent paucity of evaluation designs and useable results, heavy emphasis on summative evaluation, use of questionnaires as a predominant data collection method, addressing evaluation findings to administrators as a sole or predominant audience, reporting data in simple tables, little focus on program outcomes, and lack of effort to use evaluation findings for policy development or program improvement.

Promising practices were noted in some reports studied. These included use of: formative evaluation, multiple data-gathering methods and multiple data sources, multiple data analysis techniques, varied reporting formats, focus on multiple key program areas, and implementation of plans and strategies designed to ensure the use of evaluation findings in making positive program change.

CHAPTER 4: Case Studies in Program Evaluation Utilization in Gifted Programs

Rationale

Using results of the literature review and coding and analysis study, we were able to complete a series of recommendations on practices which *should* enhance the use of findings from gifted program evaluations for program improvement. Several evaluations included in the coding and analysis study described less exemplary practices (as defined by the *Standards*) that could inhibit the use of evaluation information. These earlier steps in the comprehensive study prepared us to look for specific factors in evaluation of gifted programs that would lead to greater evaluation utilization. Building on the literature review and the trends study, we conducted a cross case study and analysis of a purposive sample evaluation of gifted programs in our files. Our focus was on identifying evaluation designs or particular characteristics of evaluation designs and reporting strategies which yielded evidence perceived as useful by decision-makers in program development and modification.

Method

Definition

For purposes of this study, evaluation utility was defined as use of formative and/or summative evaluation information to affect a program for gifted learners in action, decision-making, or thinking about the program.

Data Gathering

Program evaluations selected for study were identified from the National Repository for Instruments and Strategies established by The National Research Center on the Gifted and Talented, (NRC/GT). We identified six school districts whose evaluation reports were most exemplary, and six whose reports were least exemplary, based on the Joint Committee on Standards for Educational Evaluation (1981).

Selection was made by coding each evaluation report according to variables such as evaluation design, method, and utility. Because of the focus of the current study on evaluation utility, a first sort of reports was completed on (1) those reports giving no recommendations for program change; (2) those giving recommendations of, but with no other attention to utility standards; and (3) those going beyond recommendations toward implementation by forming committees, developing policies, or implementing suggested changes. Those giving no recommendations were considered examples of poor practice regarding evaluation utility, while those going beyond recommendations toward implementation were considered examples of best practice. We did not know at this sorting whether oral reporting had gone beyond the written reporting.

A second sort within these two categories of reports was conducted according to evaluation audiences, applying the standard that broader disseminations are more useful than narrower ones. These reports were then arranged in chronological order. The six most recently conducted evaluations in each of the "best" and "worst" categories were given preference for study based on the pragmatic conclusion that the more recent the evaluation, the more valuable it would be in conducting a case study because of the likelihood that key personnel involved in the evaluation and subsequent decision-making process would still be available for interviews, and that their recollection of events would be more complete.

Twelve districts were selected for study and represented diversity in geography (mid-Atlantic, Northeast, Midwest, West Coast), size (ranging from districts of only three schools to districts with over 10,000 identified gifted learners), and program design (including differentiation in the regular classroom, pull-out programs, schools within schools, separate classes, schoolwide enrichment models, or combinations of delivery systems).

Initial contact for this study was made by sending letters from The National Research Center on the Gifted and Talented, University of Virginia, to school superintendents and contact persons in the 12 selected school districts asking for cooperation in the study all agreed to participate. Phone calls were then made to the district contact persons to determine appropriate informants and arrange for initial interviews. Additional informants were identified from evaluation reports or by initial interviewees as the study progressed. In a few school districts, only one individual was available. In most, between two and seven interviewees participated.

Telephone interviews were conducted in two phases. Initially, interviewers used a three-question interview protocol inquiring generally about the evaluation process and its outcome, how the process affected thinking of district personnel about the program, and how evaluation information was used. A second round of interviews (see Appendix I) followed with questions derived from the utility standards in the *Standards for Evaluations of Educational Programs, Projects, and Materials* (Joint Committee on Standards for Educational Evaluation, 1981).

Three researchers each interviewed persons from four school districts. One interviewed personnel from four of the "best" districts and one interviewed persons from the "worst" districts. A third researcher was blind to the best/worst labeling in order to serve as a check on the method used to rate districts. This researcher interviewed two districts from the "best" and two from the "worst" districts. In order to keep the third interviewer truly blind to the categories, the six best and six worst districts had been ranked within their groupings, and the bottom two of the "best" group and top two of the "worst" group was assigned to this interviewer. In effect, this interviewer was given a "middle" category—the "worst of the best" and the "best of the worst."

Data Analysis

The evaluations were studied in terms of their effectiveness in providing accurate, useful, timely, and important information for policy development. In interviews with various individuals involved with the 12 evaluations we attempted to answer the following questions:

- Were the evaluations perceived as process or product oriented?
- Did the evaluations provide useful formative and/or summative data?
- Were recommendations made to change the program in any way?
- If recommendations were made, what were the recommendations for change?
- Which changes were implemented as recommended?
- How long did program alterations remain in place?
- If no recommendations for change were made, what specific recommendations were made to maintain particular aspects of the program?
- Which program or project components continued as the recommendations suggested?

- If recommendations for retaining specific program components were made, how long did those program components remain in place?
- Were there evaluation strategies or designs, types or sources of information, data collection strategies, or instrumentation which distinguished evaluations that were influential in bringing about changes or influencing continuation of current practice or policy?
- What were the reasons for program change when and if it occurred?
- What were the perceptions of administrators and staff of the evaluations, accuracy of information, soundness of conclusions and recommendations, timeliness of presentation of recommendations, etc.?
- What factors distinguished evaluations used as formalities from those which provided data leading to program change or policy change or development?

As data were collected, summaries of each telephone interview were sent to the informant for verification or modification as necessary. Following all interviews and member checks, content analysis of interviews was conducted, with an informant's complete interview serving as a coding unit, and using pre-ordinate and emergent categories. Pre-ordinate categories included: (1) factors suggested by the literature as impacting use of evaluation findings, and (2) factors suggested to be important in the first two phases of the study as referenced earlier. Emergent categories were those which were repeated within and among interviews within the "best," "middle," and "worst" evaluation practices (e.g. informal evaluation, committee involvement, changes recommended, and changes made). Information was aggregated first for each of the 12 districts, for each of the categories (strong, middle, weak) separately, and then across categories for purposes of comparisons among them.

Triangulation of sources was obtained by interviewing several people in each school district and by interviewing several districts in each of the categories (strong, middle, weak). Triangulation of method was established by conducting a review of the districts' evaluation documents and comparing documents to the interviews and to each other. Triangulation is a "coming together" of data with each source and/or method affirming the information provided by the other sources/methods.

Results

This study confirmed many of the findings from the trends study and supported the information gleaned from the literature review. The results for the qualitative study are reported below.

Looking at the Group Characteristics

Perhaps the most critical commonality across the groups was their use of evaluation information. *All 12 districts used the information gathered through evaluation to bring about some level of change in programming.* It cannot, therefore, be concluded that evaluation utility was absent in the weaker districts and present in the stronger ones. In fact, what the study revealed was a continuum of evaluation processes and procedures, yielding a continuum of results.

The "middle group" did, indeed, serve as a check and verification that the sorting process described earlier delineated districts with weaker evaluation plans, differing in marked ways from districts with stronger evaluation plans. That is, the "best of the worst" group produced a profile much more like that of the weaker group than of the stronger; while the "worst of the best" group appeared more like the stronger group than the weaker.

Yet, the middle group did demonstrate a "middle of the road" profile when compared as a unit with the other two groups. Perhaps coincidentally and perhaps not, the two districts nearest the middle of the 12 "exchanged positions" during the course of the study. This phenomenon will be discussed later.

There were some fundamental areas of similarity across all 12 districts studied. Although it was easier to locate key personnel and an abundance of shared information was clearly more common among districts with evaluations classified as stronger, all 12 showed an interest in evaluation of gifted programs as indicated by their submission of evaluation reports and their willingness to participate in the interview process. All 12 districts did have some sort of plan to evaluate gifted programs. Thus while the reports and procedures are discussed in terms of "weak" and "strong," even the "weak" districts are more sophisticated than districts that have no systematic intent to evaluate and/or no plan for doing so.

The Continua Described

Even the evaluations were classified by utility standards when we found that this sort was easily related to strengths in the other standards.

Evaluation Focus

While each evaluation along the continuum of 12 did have a focus or purpose in its execution, it was evident that districts judged to have weaker evaluation reports had a more general focus, while the evaluations in districts at the stronger end of the continuum were characterized as having a sharper or more specific focus. For example, districts using weaker evaluation practices tended to evaluate in order to assess how one or more groups of people *felt* about the program. Districts using stronger evaluation practices, while they may have elicited constituent opinions regarding programs, also looked at more focused topics such as implementation of IEPs (individualized education plans), dropout rates among identified gifted high school students, analysis of types of services offered to gifted students, achievement compared to aptitude among gifted students, or comparison of gifted student performance with other students in a district by gender, grade level, ethnicity, and type of services received.

Participants in the Process

All school districts studied involved a variety of participants in the evaluation process. In this category, districts with plans judged stronger again differed from those with plans judged weaker in degree, and this time in two ways. First, whereas weaker evaluations tended to include data from only one or two groups of respondents (such as students or parents who completed a survey), stronger evaluations included data from multiple groups of respondents. Second, committees conducting the evaluations in districts using stronger evaluation practices tended to include informants from among groups such as students, parents, specialist teachers, general faculty, administrators, community members, and school board members. These same districts tended to report their findings to a broader audience as well. It was also the case that only the stronger evaluations involved school board members as stakeholders, rather than viewing them only as an audience to receive findings at the end of the process. Stronger evaluations were more likely to create varied channels for stakeholder input and to involve stakeholders throughout the evaluation process in order to keep them apprised of the evolving process and its findings, and to lend credibility to evaluation results.

Methods of Evaluation and Data Analysis

It is in this area where the continuum is longest, or marks the greatest difference in the extremes. Weaker evaluations tended to utilize only a form of process evaluation—that is, monitoring to determine whether programs seem to be working as people perceive they should. Even here, there is a generality among questions which speaks of a sense of how things "should be" without careful reference to program goals or documents, and a general reliance on surveys as data sources. By contrast, the stronger evaluations tended to base process evaluation upon some combination of program documents, district records, and descriptions of program practices. Further, they planned for specific comparisons between and among goals of various district programs for the gifted, and gathered data through surveys, focus groups, and interviews.

Stronger evaluations characteristically included outcome data, or findings which indicated the degree of impact of programs on student achievement through the use of such outcome assessments as achievement scores, grades, and teacher ratings of student progress. In regard to data analysis, the weaker evaluations tended to report only descriptive statistics such as tallies or listings of responses and percentages of responses. By contrast, evaluations ranked higher on utility also employed more complex descriptive statistics (such as means), inferential statistics (such as chi-square and ANOVA), and more sophisticated qualitative content analysis. These evaluations were more likely to use both qualitative and quantitative data analysis.

Implementation Plans

Once again, while all districts studied "did *something*" with the results of evaluations and were able to use them to prompt some sort of program change, the process of implementation was much more informal among the districts with weaker evaluations and much more formal and institutionalized among the others. For example, districts with weaker evaluations might encourage conversation among key staff of the gifted program regarding findings. The districts with stronger evaluations had specific, and often multi-faceted, implementation phases delineated in their evaluation plans and evident in their practices, as school officials described them in retrospect. Generally, key stakeholders were responsible for formulating the implementation plan, with the evaluator acting somewhat as a facilitator, if involved at all. In all of these districts, there was a clear expectation that implementation would occur. For example, in one district, a priority action plan is routinely developed as part of the self-study/validation process. In two other districts, recommendations are made and implementation monitored in subsequent evaluations. A fourth district conducted a self-study and invited a validation team to verify the findings of the study. Thus, in all of the districts representing strong evaluations, utilization of evaluation information was expected and provided for within the evaluation process.

Evaluation Reports

While most districts issued some sort of evaluation report, those responsible for reporting on weaker evaluations tended to share the outcomes with fewer audiences and according to a less well-defined format than did those sharing the results of stronger evaluations. (Because variety of a diverse identification was one of the variables on which the reports were initially sorted, this result was predetermined by the classification process.) Personnel involved with weaker evaluations sometimes communicated evaluation findings through informal memos to "relevant staff," or presented brief summaries of findings to the school board "in person or in writing depending on their agenda." In contrast, reporting of stronger evaluations followed a format that included evaluation purpose and concerns, evaluation method, results, findings and recommendations, and a utility or implementation section. Further, evaluation information was provided to all identified audiences (except students who presumably could have been informed through parents) via a full, formal report, an executive summary, presentations, and/or newsletters.

Purposes for Evaluations

In all districts studied, there was some political force driving the evaluation of gifted programs. Once again, the force seemed a clearer or more potent one for the districts where evaluations were stronger when compared to the others. The motivating political forces for program evaluation included parent complaints which prompted review of a program, state funding which required evaluation, and a district mandate for a five-year self-study/validation for all district programs. At the low end of the continuum were evaluations in districts whose program administrators conducted evaluations because it was in their job descriptions to do so. It was often the case in these districts that economic shortfalls would impede or diminish evaluation plans. For example, one district had not evaluated gifted programs in two years because of budget cuts, another had to relinquish use of computer assistance in data analysis because of budget constraints, and a third district had lost most of the personnel once charged with evaluation of programs for the gifted. A school board member in one district summarized the Catch-22 that typified these districts when she said, "I'm afraid we tend to work by procedure here rather than by policy, but with the current board and current financial constraints, it's not a good time to strengthen policy. It's a time when the program will probably lay low." In these districts, there was often either an implicit or an explicit fear that "talking about the program" publicly as a result of evaluation was touchy, and a decision to be made carefully, lest calling attention to the program backfire and damage it.

By contrast, because evaluation was a policy expectation rather than a procedural option in the districts where stronger evaluations had occurred, funding was not as likely to be an issue, public dialogue stemming from evaluation was standard operating procedure for many programs, and it was expected that both strengths and weaknesses would be uncovered and dealt with in a prescribed manner as a normal part of the growth process.

Qualifications of Program Personnel

Relating to purpose of evaluation was an issue of personnel training. There were two factors relating to staff training which affected the evaluations in the 12 districts investigated. Districts for which weaker evaluations were produced might (or might not) have a staff member well-trained in gifted education. They were less likely also to have personnel in the gifted program highly trained in evaluation, or at least less likely to have on-going alliances between experts in the two fields. When asked to respond to questions about determining qualifications of those who conducted evaluations, personnel in those districts said training was not an issue, or that it was not discussed in evaluation planning. In the districts characterized as having strong evaluations, informants tended to note advanced credentials of program personnel in both gifted education and program evaluation, often simultaneously present in several key persons involved in the evaluation process. It is not surprising, of course, that these districts tended to have more elaborate and sophisticated evaluation designs and procedures.

When key informants in school districts saw evaluation merely as a task to be completed as prescribed, they were more often in districts with weak evaluations; whereas their counterparts in the school districts with strong evaluations were often passionate about the power of evaluation to evoke change at both local and state levels and discussed it as a tool of choice to be used in promoting program strength.

Evaluation procedures judged to be stronger were more likely than those classified as weaker, at least occasionally, to employ external evaluators and were more likely to have findings of internal evaluations validated by someone other than the evaluator. By contrast, in weaker evaluations only one or two internal persons constructed evaluation instruments, disseminated them, analyzed and interpreted data, and promulgated findings.

This study did not support the findings of Braskamp, Brown, and Newman (1981) that readers were less likely to agree with reports written by females as opposed to males, or by evaluators or content specialists as opposed to researchers. Reports resulting in positive program change were conducted and/or written by males and females, and by program administrators (or teams) as well as evaluators.

Nature of Change Resulting From the Evaluation Process

It is important to note again that even the evaluations categorized as "weaker" in the study involved some staff member(s) who felt responsible for evaluating programs for the gifted, followed some procedure(s) for evaluation, examined evaluation findings, and as a result brought about positive program change because of what was learned.

Weaker evaluations reported changes stemming from the evaluation process such as: "Students felt the Great Books Program was boring. After discussion, we added critical thinking to this class. The students have enjoyed the class much more." "Students did not know what was required in the home classroom because they are pulled out and bused to the program . . . [so] we changed the time they returned to class to allow more contact time with the home school teachers." "Evaluations helped us realize a need to bring in more resources from the community to assist students in the program rather than assuming the g/t teacher could be all things to all students."

These are practice-specific modifications that focused directly on classroom procedures. In other instances, however, informants describing the impact of these evaluations reported changes with a more programmatic impact. "Students told us they wanted more math, and so we now have a full-time pullout program for grade 6, pre-algebra for grade 7, algebra I and II for grade 8, and a half-year algebra course with special topics."

Informants from districts with stronger evaluation reports likewise reported both practice-specific changes, "We have begun writing lessons in Spanish for identified Spanish-speaking students rather than translating English lessons into Spanish for them. "We have broadened our identification system to reflect the increasing ethnic diversity of our schools." [or] "The IEP paperwork burden which was previously overwhelming for teachers has been streamlined by the gifted coordinator." The tighter focus of evaluations in these districts is seen in reported changes such as facilitating more realistic reporting of a previously erroneously reported dropout rate, and securing program support as a result of finding that gifted learners were faring poorly when their achievement/aptitude profiles were compared with those of almost any other ability group in the district.

Profiles of the Districts

It is useful to amalgamate data gathered from districts at either end of the continuum studied in order to construct profiles of typical districts. Doing so enables comparison of the full impact of the evaluation process in weaker and stronger settings.

Profile of the Evaluation Process in a District With an Evaluation Characterized as Weaker

The coordinator of programs for the gifted in the school district may be new to her job, and the current program for gifted students may be new as well. She wants to know "whether the program works," and in addition, she has a sense that she is accountable for what is happening in the program. This will require some sort of documentation, probably an evaluation. A procedure will evolve, but not a strong policy of evaluation. "Lack of support and funding (for conducting the evaluation) are real problems."

There seem to be two approaches to deciding what to do next—either "repeating the same process as last year," or "winging it." Feeling that it would be better for several individuals to be involved in the process, the coordinator "forms a committee." "Committee members include representatives of teachers of the gifted, coordinators, principals," and perhaps parents or school board members. After several meetings with committee members, questionnaires are developed "to address concerns." Most are Likert-like surveys "with a few open-ended questions." It is perceived to be advantageous if the form is short and the questions few. "Questionnaires are distributed to cooperating teachers, students, and parents."

The coordinator herself distributes the surveys, collects them, and analyzes results by "tabulating frequencies and percentages, and noting every comment that was made." Within a month or two of administering the survey, the coordinator shares "results with committee members for discussion about recommendations on program improvement or development." "The information is then shared with the superintendent who, in turn, informs the school board of the additional opportunities for students."

Profile of the Evaluation Process in a District With an Evaluation Characterized as Stronger

In this district, the coordinator of gifted programs has been in her current position for some time. She is aware of the political mandate for evaluation which exists in her district for programs for the gifted as it does "for all other programs with a curriculum." There is a policy that both requires and supports evaluation. She also understands the power of evaluation to improve the program and "to build awareness of and support for what we are doing." "We work hard to look at ourselves honestly," she says. "We realize when we need to change, and that is healthy." "Politically, evaluation findings allow support to be built for programs."

Here, evaluation is an on-going and multi-faceted process. "There is formative evaluation of everything specialists do in the classroom with general teachers." "The teachers tell us what is working and what we can modify. In the process, they also come to understand our goals better, too." And there are feedback sheets on "how teachers feel about administration of the testing program we are in charge of to assist us with the management of testing." "We are very diligent in following through with findings." "There is at least one kind of survey every semester—periodic surveys of building principals, students, and teachers in that school." There are "standard, self-monitoring devices in place in schools" and staff there with enduring responsibility for interpreting findings to building personnel as they relate to that school.

There is a team of district professionals who can collaborate on evaluation procedures—at times members of the gifted education staff with strong credentials in evaluation as well, at times a partnership between a district evaluation department and members of the gifted education staff. While one person assumes responsibility for the evaluation process as it relates to gifted education, it is a leadership responsibility, and not sole responsibility. There is a steering committee for gifted programs which plays a key role in evaluation, but there are other groups and committees engaged in the process as well. "We don't want to rely just on one source."

There is also a strong awareness of the varied stakeholders in the district. Stakeholders are a part of evaluation planning, execution, and follow-up. These committees assist in determining specific program areas to be studied and propose questions whose answers could be valuable in providing program support. "We want them to have all the information they need." "To understand what we are about." "To keep them apprised of findings so there are no surprises in the end." "So they will buy into the evaluation." "So they support program changes which follow." When findings are generated, they are brought back to stakeholder committees "first orally, and then in preliminary reports." "To give the stakeholders a chance to see whether the findings made sense and to determine if the recommendations are feasible."

In addition to process evaluation, the district examines outcome indicators. "The school board pays some attention to achievement data." "Recently, we conducted a panel study comparing test data for all students. In our self-contained program, all scores went up, which is amazing given the likelihood of regression to the mean. There was also strong evidence that these programs were benefiting minority achievement." "We have begun using portfolios as a means of assessing the impact of the critical and creative thinking components in our program."

From time to time, external evaluations of the program are conducted. "There is a built-in suspicion that if the g/t staff is conducting all the evaluations, they can't be really legitimate." "A few years ago there was a huge external evaluation with university support to set a future direction for our gifted programs. The process was useful and we have built steadily on its findings."

Data analysis is done with appropriate technical support and qualitative and/or quantitative methods appropriate to the questions asked and evaluation formats used. A final, formal report is released, on a pre-set time-line, to appropriate groups including stakeholders, school board, staff, and frequently with report summaries available for new media and parent groups. The formal report is written in a format similar to that of a research study, with appropriate data tables and accompanying explanations. A standard part of the report is an implementation section, "outlining what is to be done as a result of the evaluation findings, who has oversight responsibility for the new plans, and a time-line for completion." There is also a plan in place "to monitor next year how we've done with our commitment."

A Cross-Group Comparison

The great difference emerging between those school districts categorized as having weaker evaluation plans and those having stronger ones lies in sharply contrasting levels of training and of support. There is the intent to evaluate and to do it to the best of one's capacity in both settings—and, in fact, there are indications of success in both groups as measured by positive program changes that arise from evaluation findings.

In the settings from which stronger evaluations emanated, those in charge of the evaluation process understand evaluation as a field of study. They use vocabulary like "stakeholders," "formative evaluation," "outcome indicators," and "chi-square." They understand the peculiar pitfalls of measuring academic growth in students who top-out on tests, and can discuss the use of portfolios, comparison of achievement and aptitude scores, and regression to the mean. They have a level of political sophistication that helps them see both a need and a means for building networks of support through evaluation processes for the programs they administer. Further, they have access to technical and collegial support in the evaluation process, a reality which further enhances the range and potency of the evaluation process.

By contrast, coordinators in the districts categorized as having weaker evaluations sense a need to know "how things are going," and they use the only tool at their disposal—common sense. They work alone (or perceive that they do), and join forces with others via committee, gaining a sense of partnership, and feeling reinforced in their common sense strategies.

A Tale From the Middle Group

It was at least symbolic that the two districts directly in the middle of the ranking of 12 "changed places" as the study unfolded. The district whose evaluation ranked as "strong among the weak," had clearly moved up in the world since its original materials had been received. A new coordinator had come aboard—one who used terms like "portfolio assessment" and "outcome-based evaluation." She was moving away from sole use of attitude surveys. "We need to look at performance and program benefits in achievement instead of just whether parents, students, and teachers like the program." She has used the drawings of primary students to study attitude changes about science and scientists in youngsters who have participated in a magnet program where they work directly with scientists, compared with youngsters who have not had that opportunity. She was working to integrate some evaluation components of services for gifted learners into the evaluation processes of individual schools. Furthermore, she talked about working with other administrators and board members, as well as using the evaluation data which shows a gap "between predicted and actual test scores of gifted students for action at both local and state levels."

In the district whose evaluation report was initially classified as "weakest among the strong," there was a clear backslide. In this setting, there had once been a coordinator of gifted programs who worked with a strong and knowledgeable planning committee on the district-mandated evaluation process. Two people who worked on the committee had Ph.D.s in evaluation, and the other was working on a Ph.D. "There were also consultants involved in developing the evaluation processes and procedures." From both oral reports and evaluation documents, the evaluation system was effective in bringing about program improvement.

At some point, staff assignments changed, and the new coordinator (who was assigned only a small portion of her time for administering gifted programs) inherited and elected to maintain the previous evaluation design. Talking about the plan, she explained that she "wasn't quite sure how decisions were made regarding questions to be asked in the evaluation process." "The chief audience for the evaluation findings was the Gifted and Talented Planning Committee." "Principals were also given results of the evaluation by schools and helped to analyze them." "Principals who had preconceptions probably didn't change as a result of the meetings, but those who were open to suggestions and wanted to listen were helped to make changes." "Ultimately these meetings were instrumental in leading to a model shift in the district's gifted program." "There was no systematic follow-up on these meetings to see whether plans had been executed."

At this point, the "new" coordinator has moved on. A new program has been put in place "based on evaluation findings." The "school board has adopted the new program, but not funded it." "There is no evaluation procedure in place for the new program . . . and there is no staff to work on evaluation." "Regular classroom teachers are supposed to assume responsibility for the [new] model as well as their own assignments. It makes their attitude toward the program negative. There is no acknowledgment of what they are doing."

Decisive Factors in Use of Findings

This study indicates two key factors which promote use of evaluation findings in districts studied—will and skill. It appears that the will to evaluate on the part of some key personnel in a district, supplemented with systematic procedures for doing so, results in generation of evaluation findings and translation of those findings into program change. This will to evaluate existed in all the school districts studied.

The second factor—skill in evaluation and related processes—appears the demarcation between the two categories of evaluations and affects the robustness of program change stemming from evaluation findings. Utilization appeared more likely and changes from the findings more potent and systemic in direct relationship to the following conditions:

1. Evaluation of gifted programs was a part of a district-wide policy requiring routine evaluation for all program areas.
2. Systematic written plans were in place delineating steps and procedures for ensuring implementation of findings.
3. Multiple stakeholders were consistently involved in planning, monitoring, and reviewing the evaluation process and its findings.
4. Stakeholders played an active role in planning for and advocating before policy makers for program change based on evaluation findings.
5. Key program personnel were knowledgeable about gifted education, evaluation, the political processes in their districts, and the interconnectedness of the three.

Concerns

Perhaps the major concern highlighted by this project is the paucity of evaluation reports/results made available to the NRC/GT. This is likely the result of lack of gifted program evaluations or dissatisfaction with evaluation designs and results. These explanations are considered more likely in light of the high number of responses received at the NRC/GT during the same time frame with regard to identification policies and instruments.

Another concern is that evaluations that are carried on tend to be summative evaluations addressed to administrators, dealing with concerns raised by administrators, with information often gained through questionnaires as the sole method of data collection. Further, information from these evaluations is often disseminated in the form of simple data tables, with little focus on program outcomes. Such evaluation tends not to be associated with efforts to use the information for policy development or program improvement. Seeley (1986) aptly described such evaluation as "academic gymnastics" (p. 286).

Further, where external evaluators are used, they often focus their reports on the needs of the research community rather than on those of the client. Often a research paradigm is employed that ignores recent thinking about evaluation design and utilization.

Promising Practices

A number of promising practices seem to be emerging in the evaluation of gifted programs. First of all, a large subset of the evaluations analyzed in this report employed a formative type of evaluation. Their expressed intent was program improvement. Further, many of the evaluations studied incorporated multiple data-gathering methods from multiple data sources; many used multiple data analysis techniques; and a number reported results through multiple formats. This is important given the apparent association of the use of multiple methods, sources, analysis techniques, and reporting formats with utility practices that produce policy development and program improvement.

Second, in accord with Callahan (1986) and Carter and Hamilton (1985), many of the evaluations focused on a number of key areas in the gifted program rather than settling for generalized impressions about the program. While evaluation of key program components tended to be subjective in nature, important programming issues were dealt with across multiple audiences.

Finally, the importance of making evaluation information useful appears to be taking root. Most evaluations at least generated recommendations, and many went beyond recommendations to formulate committees, goals, action plans, and policies.

CHAPTER 5: Conclusions and Summary

Implicit in conducting an evaluation are the assumption that appropriate instruments/data collection strategies will be used, that evaluations are designed to incorporate standards of ethical and sound evaluation practice, and that there is an intention to *use* evaluation findings in some way. We begin an evaluation with the expectation that evaluation findings will be helpful in directing the thinking of program planners or in creating a road map for action. For our hopes to be fulfilled, however, evaluation findings must be acted upon by one person, or many. Unfortunately, it is often the case that evaluation findings are not used, resulting in wasted effort and cost as well as loss to students if potential program improvements are not made.

Those who seek educational improvement through evaluation thus need to have information about appropriate instruments, the interactions of variables in evaluation designs, and factors that promote or inhibit the use of evaluation findings. Within this technical report we have provided information on a collection of instruments used in the evaluation of gifted programs and an instrument for assessing the technical properties of those instruments (see Appendix J). Second, we provided a review of the literature on increasing the utility of evaluations. Third, we provided an analysis of current evaluation reports on factors which characterize current evaluation reports. Finally, we studied the characteristics of schools where evaluations were characterized as meeting the criteria of the standards for evaluation and those which did for information on utilization and factors which made these evaluations come about and succeed. The studies of evaluation utilization, combined with a study of particular evaluation needs of gifted education offer direction in planning and conducting "useful" evaluations of gifted programs.

Increasing Use of Evaluation Results in General: The Literature

The Impact of Economics, Politics, Definition, and Design

Experts in the field of evaluation suggest a number of factors that improve the likelihood that the results of any evaluation are useful, and therefore used.

1. Begin with funds and commitments. While it is difficult for evaluators to control the economic and political situations that surround them, it is important to note that evaluation results are less likely to be used or to be used appropriately if there are no funds to implement recommendations. Further, if there is a lack of commitment to the program or to program change on the part of people in positions of power and influence, little attention will be given to evaluation findings. While such economic and political realities are difficult to eradicate, it may be that other factors under the evaluator's control can positively influence these realities.
2. Select clear, appropriate designs. Within evaluator control are several other factors to which evaluators should attend. It is important to plan evaluations from the earliest stages of program planning, to define the purposes of the evaluation, and to select an evaluation design appropriate to the program and program features which will be evaluated. For example, quantitative (statistically oriented) designs may be especially useful when outcomes are a focus. However, qualitative (descriptive and case study in orientation) designs are more appropriate when processes within a program are studied or when complex settings are examined. A combination of qualitative and

quantitative designs is called for when both processes and outcomes are of concern (Carter & Hamilton, 1985; Smith, 1980;).

The Impact of Message Source, Content, and Receiver

The work of Braskamp, Brown, and Newman (1981) suggests that variables which affect evaluation utility can be grouped as message source, message content, and message receiver. In other words, how will the evaluator, the evaluation report, and the audience itself impact use of evaluation findings?

1. Establish credibility of evaluator and evaluation process. With regard to message source or the evaluator, is important that the evaluator be credible to those who will receive the evaluation report and that the evaluator carefully explain procedures and rationales used in determining findings and recommendations (Joint Committee on Standards for Educational Evaluation, 1981).
2. Prepare understandable and well-documented, but succinct reports. Message content has to do with the report itself. Information collected should be of sufficient breadth and collected in ways which allow pertinent questions to be pursued and in ways which address the needs of a variety of appropriate audiences (Joint Committee on Standards for Educational Evaluation, 1981). Using multiple data gathering methods (e.g., surveys, observations, interviews, and standardized measures) increase the usefulness of findings, as does drawing upon a variety of data sources (e.g., students, teachers, parents, school board members, administrators). Reports which are timely and free of jargon and masses of data are typically more useful as well (Bickel & Cooley, 1981; Kennedy, Apling, & Neumann, 1980; King, Thompson, & Pechman, 1981).
3. Direct reports to appropriate audiences at appropriate times. An examination of data relating to receiver or audience characteristics leads to the conclusion that it is important to clearly identify clients and audiences of the evaluation, and to involve them actively throughout the evaluation design, data collection, and data analysis. People who feel a clear need for evaluation are more likely to utilize findings than those who do not. Effective and on-going communication with clients and audiences is important in establishing a sense of the worth of the evaluation. Similarly, it is important that the evaluation report be disseminated to clients and relevant audiences in a timely fashion which allows information to be received while it is useful and can be acted upon (Bickel & Cooley, 1981; D'Amico & Dawson, 1985; Kennedy, Apling, & Newman, 1980).

Special Challenges in Evaluating Gifted Programs

Programs for gifted learners are marked by certain complicating characteristics which must be understood and accounted for in the planning and execution of evaluations so that results are likely to be used. Some of the problems posed in assessing the effectiveness of gifted programs relate to the design or articulation of the programs themselves, others to issues of evaluation design and measurement. Suggestions which emerged for dealing with these issues include:

1. Clearly delineate program goals. Callahan (1983) points out that gifted programs often suffer from poorly delineated program goals. In instances where program goals are unstated, vague or unfocused, it is difficult to design an evaluation that addresses the impact of the program. Further, goals of programs for gifted learners are long-term ones (e.g., development of creative or critical thinking skills, development of skills of independent learning) and are inappropriately assessed by measures better suited to demonstrating short-term change (e.g., mastery of information).
2. Carefully address design and measurement issues. Many of the confounding traits of programs for gifted learners have an impact on measurement and design decisions within the evaluation. For example, goals of gifted programs are likely to be holistic, complex, product-oriented, and individualistic, thus poorly measured by standard means which focus on group goals and norms and behavioral objectives which focus on goals that are simpler or at a lower level (Callahan, 1983; Ganapole, 1982; Gilberg, 1983; Renzulli, 1984). Standardized tests do not measure the sort of advanced learning which is the hallmark of strong programs for gifted learners (Callahan, 1983; Gilberg, 1983; Renzulli, 1984).

Gifted learners typically score at the top of standardized measures as part of the criteria for entering gifted education programs. It is impossible, then, to demonstrate growth by using the same or similar standardized measures of outcomes because there is no room for growth on that test scale (Callahan, 1983). Standardized tests administered at grade level have low ceilings and are thus not appropriate for assessing student growth at the top of their scales. In addition, they are typically poor at demonstrating growth in older students, creating a greater difficulty documenting growth in secondary gifted students (Renzulli, 1984).

When standardized tests are normed on heterogeneous groups, their norms are not necessarily reliable for relatively homogeneous groups, such as groups of gifted learners (Callahan, 1983; Gilberg, 1983; Renzulli, 1984).

In regard to measurement and design concerns, a number of alternatives to traditional approaches are helpful:

1. Use out-of-level tests where *valid* for the trait/outcome assessed to combat the low ceiling effect (Callahan, 1983).
2. Develop and use common criteria for examining student products and portfolios, and establish inter-rater reliability in application of the criteria (Beggs, Mouw, & Barton, 1989).
3. As alternatives to randomized experiments, consider use of carefully matched groups between schools, one receiving the intervention to be assessed, one not receiving it (Carter, 1986; Payne & Brown, 1982). Or consider a time-series design in which all groups of gifted learners receive the target intervention, but at various times, thus serving as controls for one another (Callahan, 1983). Another alternative is retrospective pretesting in which students receive an intervention, take a test or survey which assesses post-intervention knowledge or opinions, then take the same test or survey which asks them how they would have answered the questions prior to the intervention. Students are thus giving their own sense of how their knowledge or feelings have changed as a result of the intervention being studied, and the data can be used to compare mean differences (Payne & Brown, 1982; Carter, 1986). A contrast group (rather than a control group) in which an existing group or to-be-generated data set serves as a contrast to

results from the intervention in question may serve the evaluation function. Use of a contrast group rather than a more traditional control group acknowledges the fact that even random assignment of students to experimental and control groups cannot eliminate factors which call into question the cause of findings. "Control" is often difficult to achieve in educational evaluation, and using a contrast group acknowledges that fact while it appropriately separates evaluation studies from experimental studies (Payne & Brown, 1982). Finally, it may be useful to target intervention in both regular and gifted/talented classes to measure the breadth and depth of achievement and rate of learning of the two groups in order to better understand differentiated education (Payne & Brown, 1982).

Experimental designs raise issues of withholding services from some qualified students, right to knowledge of treatment, and the John Henry effect which may occur when a non-treatment group in an experiment reacts with the intent to demonstrate that they are equally skilled or able in the area being measured (Callahan, 1983; Payne & Brown, 1982).

The Messenger, the Receiver, and the Evaluations for Gifted Programs

While the literature of evaluation utilization in gifted education is limited, a few writers and researchers do address issues related to message source, message content, and receiver characteristics as these factors relate to increasing the usefulness of evaluation findings in programs for the gifted. They suggest:

1. Prepare staff carefully for the evaluation. In regard to message source, Gilberg (1983) encourages us to ensure that both staff and evaluators are trained to carry out and analyze the results of the evaluation. Ganapole (1982) specifies the need to prepare and describe rules of scoring prior to administration of tests in evaluating gifted programs.
2. Address questions important to the evaluation audiences. In writing about message content and gifted programs, Callahan (1986) reminds us to address the needs of both internal and external audiences of programs, and to address questions helpful in making decisions that can have an impact on program quality. Such questions may address the function, components, goals, activities, and structure of the program in question. Further, questions may relate to program areas that are of central importance, potential problems in the program, level of resources, undesirable change brought about by the program, conflict with values of other stakeholders, loss of power, inconsistency between program goals and implementation of those goals, lack of understanding of goals, and personal bias. She also reminds us that evaluation questions should be specific to the program being evaluated, unlike research questions which seek generalizability to other settings.
3. Use a variety of data collection strategies. There is a need to use a variety of data collection modes in order to respond to the varied needs of different constituencies of gifted programs (Gilberg, 1983; Janesick, 1989; Rimm, 1982; Turner, Hartman, Nielsen, & Lombana, 1988), and a need to describe in detail the program being evaluated so that the evaluator has a clear sense of what constitutes the program and which factors impact gifted learners in specific ways (Callahan, 1983).
4. Know the biases of decision-makers. In regard to receiver characteristics which may affect utilization of evaluation results in gifted programs, it is

necessary for the evaluator to identify decision-makers clearly and to understand the actions over which they have control (Callahan, 1986; Dettmer, 1985; Gilberg, 1983; Renzulli, 1984; Rimm, 1982). Gilberg (1983) encourages evaluators to find out what courses of action will result from evaluation findings, and to make recommendations with an eye toward improving the program. Dettmer (1985) recommends that: (a) self-studies be conducted by local gifted/talented advisory councils as a result of evaluation findings, (b) specific recommendations be made as a result of the self-study, (c) reports of the self-study and recommendations be prepared for each stakeholder group, and (d) actions for carrying out the recommendations be initiated.

Summary Guidelines for Conducting Useful Evaluations of Gifted Programs

Both the general literature of evaluation utilization and the literature of gifted education provide guides which can be summarized in four general principles.

1. Make evaluation a part of program planning from the earliest stages of program development.
2. Clearly identify all audiences who have an interest in or need for evaluation results, and involve them in the evaluation process.
3. Develop evaluation designs that address complex issues of measurement in gifted programs.
4. Avoid reliance on traditional standardized measures that offer little promise of reflecting academic growth in gifted students and are involved in assessing goals for gifted learners.

In times when programs for gifted learners must compete for unusually scarce resources, it is imperative that program administrators and evaluators of gifted programs understand the need to plan and conduct evaluations that are appropriate for those programs and that facilitate use of findings for program improvement.

Increasing Evaluation Utilization: Our Studies

Where intent to evaluate gifted programs exists, some form of evaluation is likely to evolve. Even when such evaluation schemes are relatively "weak," at least in comparison to evaluation plans that closely follow utility standards such as those developed by the Joint Committee on Standards for Educational Evaluation (1981), utilization of evaluation findings can and does occur in ways that result in positive program change.

It is clear, however, that more robust evaluation designs and procedures evolve when responsible personnel have specific training in evaluation, in gifted education, and in problems of evaluating gifted programs—and when they have support in the way of well-trained colleagues and policy expectations. Such program personnel have access to vocabulary, procedures, and a level of political sophistication that enable them to maximize the capacity of evaluation both to chart program growth and amass program support, including economic support.

The example of the "middle districts" which reversed places offers a cautionary note. Evaluation procedures carry with them a certain potency—somewhat like a moving automobile. Once in motion, if they are not properly steered, their power can veer in inconvenient, if not dangerous, directions. Informed operators may plan to reach, in at least relative safety, desirable destinations. Once set in motion, a driverless vehicle, or even a vehicle manned by a novice, can imperil the passengers.

Thus while "good intentions" may yield progress in a desirable direction, the process can also go awry.

The clearest need emerging from the study is for the training of program personnel in gifted education and program evaluation, the problems gifted programs present in assessment of student growth, and in evaluation methodology appropriate for assessing such programs. Even many of the "strong" districts showed only fledgling movement in the direction of experimental design to demonstrate student growth (Beggs, Mouw, & Barton, 1989; Callahan, 1983; Carter, 1986; Payne & Brown, 1982), and few appear to have tapped the range of possibilities of qualitative design for evaluating gifted programs (Janesick, 1989, Lundsteen, 1987).

Certainly the "weaker" districts have need for personnel with knowledge of how to employ varied data collection modes (Gilberg, 1983; Janesick, 1989; Rimm, 1982), how to address concerns of both internal and external audiences by asking questions which are relevant, useful, and important and which will thus directly facilitate positive and powerful decision-making (Callahan, 1986), how to identify decision-makers at various levels as well as actions over which they have control (Callahan, 1986; Dettmer, 1985; Gilberg, 1983; Renzulli, 1984; Rimm, 1982) and how to find out what course of action will result from data supplied, as well as how to make recommendations with an eye toward program improvement (Gilberg, 1983).

To function at a lesser state is to compromise the positive possibilities of education. One interviewee expressed an added sense of urgency for such understandings. "Programs for the gifted operate under some threat because they are not valued by society as a whole. Therefore, all of our staff members know there is a need to put forth effort to achieve a high degree of improvement."

Suggestions for Improving Gifted Program Evaluation

While certain measurement and practical problems somewhat unique to gifted education make effective evaluation difficult, suggestions for overcoming the obstacles and conducting more useful evaluations can be derived from the general literature on evaluation utility and on rating the success of evaluations in the schools examined.

1. Make evaluation procedures a part of planning from the earliest stages of program development (including clear program descriptions and goals), and develop a specific plan for the use of evaluation findings.
2. Ensure that evaluators are trustworthy and knowledgeable of both gifted education and evaluation.
3. Provide adequate funding and time for appropriate evaluation procedures to be followed.
4. Clearly identify all audiences who have an interest in or need for evaluation results and involve them in the evaluation process.
5. Ask evaluation questions that are well focused to provide information about the goals, structures and activities of the program being evaluated—questions that will aid in making significant program modifications.

6. Use multiple data sources (e.g., teachers, parents, students, administrators, school board members) in order to understand the values of varied groups of stakeholders.
7. Develop or select assessment tools that address the complex issues of measurement that characterize outcomes of gifted programs.
8. Consider the use of a combination of qualitative strategies and quantitative methods as time series design, using students as their own controls, retrospective pretesting, case studies, etc.
9. Avoid reliance on traditional standardized measures that offer little promise of reflecting academic growth in gifted learners unless standardized tests measure what you value as the outcomes of your gifted program.
10. Use a variety of data gathering methods designed to reflect the unique structure and goals of programs for gifted learners (i.e., out-of-level testing, portfolio assessment, product rating with demonstrated inter-rater reliability).
11. Describe procedures for data collection and interpretation fully and in jargon-free language so that audiences understand processes that were followed and conclusions that were drawn.
12. Disseminate reports to all appropriate audiences in a timely fashion and with recommendations designed to encourage follow-through.

References

Alkin, M. (1980). Uses and users of evaluation. In E. L. Baker (Ed.), *Evaluating federal education programs* (Report No. CSE-R-153, pp. 39-52). Los Angeles: California University, Center for the Study of Evaluation. (ERIC Document Reproduction Service No. ED 205-599)

Alkin, M., Dailak, R., & White, P. (1979). *Using evaluations: Does evaluation make a difference?* Beverly Hills, CA: Sage Publications.

Alkin, M., & Stecher, B. (1981, April). *Evaluation information use by site level decision makers: Methodological issues.* Paper presented at the annual meeting of the American Educational Research Association, Los Angeles, CA. (ERIC Document Reproduction Service No. ED 202-890)

Beggs, D., Mouw, J., & Barton, J. (1989). Evaluating gifted programs: Documenting individual and programmatic outcomes. *Roeper Review, 12*(2), 73-76.

Bickel, W., & Cooley, W. (1981). *The utilization of a district wide needs assessment.* Pittsburgh, PA: Pittsburgh University, Learning and Research and Development Center. (ERIC Document Reproduction Service No. ED 218-339)

Bissell, J. (1979). *Program impact evaluations: An introduction for managers of Title VII projects. A draft guidebook.* Los Alamitos, CA: South West Regional Laboratory for Educational Research and Development. (ERIC Document Reproduction Service No. ED 209-301)

Braskamp, L., Brown, R., & Newman, D. (1981). Studying evaluation utilization through simulations. *Evaluation Review, 6*(1), 114-126.

Brown, F. G. (1980). *Guidelines for test use: A commentary on the standards for educational and psychological testing*. Washington, DC: National Council on Measurement in Education.

Brown, R., Newman, D., & Rivers, L. (1980). Perceived need for evaluation as an influence on evaluator's impact on decision-making. *Educational Evaluation and Policy Analysis, 2,* 67-73.

Callahan, C. M. (1983). Issues in evaluating programs for the gifted. *Gifted Child Quarterly, 27,* 3-7.

Callahan, C. M. (1984). Using evaluation results to improve programs for the gifted. *Journal for the Education of the Gifted, 7,* 60-74.

Callahan, C. M. (1986). Asking the right questions: The central issue in evaluating programs for the gifted and talented. *Gifted Child Quarterly, 30,* 38-42.

Callahan, C. M., & Caldwell, M. S. (1986). Defensible evaluations of programs for the gifted and talented. In C. J. Maker (Ed.), *Critical issues in gifted education: Defensible programs for the gifted* (pp. 277-296). Rockville, MD: Aspen.

Callahan, C. M., & Caldwell, M. S. (1993). Establishment of a National Data Bank on identification and evaluation instruments. *Journal for the Education of the Gifted, 16*, 201-219.

Callahan, C. M., Tomlinson, C. A., & Pizzat, P. M. (n.d.). *Contexts for promise: Noteworthy practices and innovations in the identification of gifted students.* Charlottesville, VA: University of Virginia, The National Research Center on the Gifted and Talented.

Carter, K., & Hamilton, W. (1985). Formative evaluation of gifted programs: A process model. *Gifted Child Quarterly, 29*, 5-11.

Carter, K. R. (1986). A cognitive outcomes study to evaluate curriculum for the gifted. *Journal for the Education of the Gifted, 10*, 41-55.

D'Amico, J., & Dawson, J. (1985). *Cooperation or coaptation? The benefits and limitations of involving program developers in program evaluation.* (ERIC Reproduction Service No. ED 262-108)

Datta, L. (1979). *O thou that bringest the tidings to lions: Reporting the findings of educational evaluations.* Paper presented at the annual Johns Hopkins University National Symposium on Educational Research, Baltimore, MD.

David, J. (1981). Local uses of Title I evaluations. *Educational Evaluation and Policy Analysis, 3*, 27-39.

Dettmer, P. (1985). Gifted program scope, structure and evaluation. *Roeper Review, 7*, 146-152.

Fairweather, G. (1981). Community psychology for the 1980s and beyond. *Evaluation and Program Planning, 3*, 245-250.

Gallagher, J. J. (1988). National agenda for educating gifted students: Statement of priorities. *Exceptional Children, 55*, 107-114.

Gallagher, J. J., Weiss, P., Oglesby, K., & Thomas, T. (1983). *The status of gifted/talented education: United States survey of needs, practices, and policies.* Ventura, CA: Office of the Ventura County Superintendent of Schools.

Ganapole, S. (1982). Measuring the educational outcomes of gifted programs. *Roeper Review, 5*, 3-7.

Gilberg, J. (1983). Formative evaluation of gifted and talented programs. *Roeper Review, 6*, 43-44.

Gold, N. (1983, March). Stakeholder and program evaluation: Characterizations and reflections. *New Directions for Program Evaluation, 17*, 63-72.

Guba, E. (1978). Toward a methodology of naturalistic inquiry in educational evaluation. *CSE Monograph Series in Evaluation, 8.* Los Angeles: University of California, Center for the Study of Evaluation.

Hoepfner, R., Hemenway, J., DeMuth, J., Tenopyr, M. L., Granville, A. C., Petrosko, J. M., Krakower, J. K., Silberstein, R., & Nadeau, M. (Eds.). (1972). *CSE-RBS test evaluations: Tests of higher order cognitive, affective, and interpersonal skills.* Los Angeles: Evaluation Technologies Program, Center for the Study of Evaluation and the Humanizing Learning Program, Research for Better Schools.

Hoepfner, R., Strickland, G., Jansen, P., & Patalino, M. (Eds.). (1970). *CSE elementary school test evaluation.* Los Angeles: UCLA Graduate School of Education, Center for the Study of Evaluation.

Hunsaker, S. L., & Callahan, C. M. (1993). Evaluation of gifted programs: Current practices. *Journal for the Education of the Gifted, 16*, 190-200.

Janesick, V. (1989). *Stages of developing a qualitative evaluation plan for a regional high school of excellence in upstate New York.* Paper presented at American Evaluation Association, San Francisco, CA.

Joint Committee on Standards for Educational Evaluation. (1981). *Standards for evaluations of educational programs, projects, and materials.* New York: McGraw-Hill.

Kennedy, M., Apling, R., & Neumann, W. (1980). *The rate of evaluation and test information in public schools.* Cambridge, MA: The Huron Institute.

King, J., Thompson, B., & Pechman, E. (1981). *Improving evaluation use in local school settings* (NIE-G-80-0082). Washington, DC: National Institute of Education. (ERIC Reproduction Service No. ed. 214 998)

Lundsteen, S. (1987). Qualitative assessment of gifted education. *Gifted Child Quarterly, 31,* 25-29.

Marshall, C. (1984). The case study evaluation: A means for managing organizational and political tensions. *Evaluation and Program Planning, 7,* 253-266.

Mathis, W. (1980, April). *Evaluating: The policy implications.* Paper presented at the American Educational Research Association, Boston, MA.

Ostrander, S., Goldstein, P., & Hull, D. (1978). Toward overcoming problems in evaluation research: A beginning perspective on power. *Evaluation and Program Planning, 1,* 187-193.

Patton, M. (1988). Six honest serving men for evaluation. *Studies in Educational Evaluation, 14,* 301-330.

Patton, M. (1989). *Qualitative evaluation methods.* Newbury Park, CA: Sage Publications.

Patton, M. (1990). *Qualitative evaluation and research methods.* Newbury Park, CA: Sage Publications.

Patton, M., Grimes, P., Guthrie, K., & Brennan, J. (1977). In search of impact: An analysis of utilization of federal health evaluation research. In C. H. Weiss (Ed.), *Using social research in public policymaking* (pp. 141-163). New York: Free Press.

Payne, D., & Brown, C. (1982). The use and abuse of control groups in program evaluation. *Roeper Review, 5,* 11-14.

Raizen, S., & Rossi, P. (Eds.). (1981). *Program evaluation in education: When? How? To what ends?* Washington, DC: National Academy Press. (ERIC Document Reproduction Service No. ed. 205 614)

Renzulli, J. S. (1975). *A guidebook for evaluating programs for the gifted and talented.* Ventura, CA: Office of the Ventura County Superintendent of Schools.

Renzulli, J. S. (1984). Evaluating programs for the gifted: Four questions about the larger issues. *Gifted Education International, 2*(2), 83-87.

Rimm, S. (1982). Evaluation of gifted programs: As easy as ABC. *Roeper Review, 5,* 8-11.

Seeley, K. R. (1986). Evaluation for defensible programs for the gifted. In C. J. Maker (Ed.), *Critical issues in gifted education* (pp. 265-276). Rockville, MD: Aspen.

Smith, N. (1980). The feasibility and desirability of experimental methods in evaluation. *Evaluation and Program Planning, 3,* 251-256.

Stake, R. (1975). *Evaluating the arts in education: A responsive approach.* Columbus, OH: Merrill.

Tomlinson, C., Bland, L., & Moon, T. (1993). Evaluation utilization: A review of the literature with implications for gifted education. *Journal for the Education of the Gifted, 16,* 171-189.

Tomlinson, C., Bland, L., Moon, T., & Callahan, C. M. (1994). Case studies of evaluation utilization in gifted education. *Evaluation Practice, 15,* 153-168.

Turner, S., Hartman, J., Nielsen, L., & Lombana, J. (1988). Fostering utilization through multiple data gathering methods. *Studies in Educational Evaluation, 14,* 113-133.

Worthen, B. R., & Sanders, J. R. (1987). *Educational evaluation: Alternative approaches and practical guidelines.* New York: Longman.

Appendix A

Tables

Table A-1.

Summary of Databases on the Evaluation of Gifted Programs

Database Name	Description of Contents	Number of Entries*
EVALDES	articles related to the design of evaluation systems.	15
EVALUTIL	articles about using information from evaluation.	38
EVALLOC	contains instruments and information from local school systems about their evaluation procedures.	332
EVALPUB	published and standardized instruments used in the evaluation of gifted students and/or gifted programs.	103
EVALREPT	reports of program evaluations which have been sent to the NRC/GT by schools or school districts.	114
EVALNOST	published, nonstandardized instruments used in the evaluation of gifted students or gifted programs.	164

*as of 3/1/93

A letter was sent to all contributors of locally developed materials asking for permission to release these materials. Only materials from school districts that have given permission for distribution are included in the database used to fill requests for local instruments, although all instruments were included when analyses of the data were conducted for our report. Any local instrument released also contains the name and address of a contact person in the district which developed the instrument.

EVALDES(ign) files contain articles related to the design of evaluation systems. Particular attention is placed on information about the appropriateness of the design suggested for evaluating gifted students and/or gifted programs.

EVALUTIL(ization) files contain articles about using information from evaluations. Particular attention is placed on the relevance of the utilization strategies suggested to gifted programs.

EVALLOC files contain instruments and information from local school systems about their evaluation procedures.

EVALPUB files contain published and standardized instruments used in the evaluation of gifted students and/or gifted programs.

EVALREPT contain reports of program evaluations which have been sent to the NRC/GT by schools or school districts. They are reviewed for the basic features of the process: methodology, analysis, intended audience, etc.

EVALNOST files contain published, nonstandardized instruments used in the evaluation of gifted students or gifted programs.

Table A-2[*]

Frequencies of Evaluation Types

Evaluation Type	f	%
Summative	39	55.7
Formative	25	35.7
Combined	2	2.9
Needs Assessment	2	2.9
Other	5	7.1

Table A-3

Frequencies of Evaluation Models

Evaluation Model	f	%
Management Centered	40	57.1
Objectives Centered	20	28.6
Product Centered	10	14.3
Participant Centered	4	5.7
Combined	3	4.3

Table A-4

Frequencies of Evaluator Types

Evaluator Type	f	%
Internal	41	58.6
External	30	42.9
Combined	1	1.4

---

[*] For each of the Tables A-2–26, frequencies may add to more than 70 and percentages to more than 100 because of multiple categorizations.

Table A-5

Frequencies of Data-Gathering Methods

Data-Gathering	f	%
Questionnaire	54	77.1
Test	26	37.1
Document Analysis	23	32.9
Observation	22	31.4
Interview	21	30.0
Meeting	8	11.4
Other	5	7.1
Multiple	43	61.4

Table A-6

Frequencies of Data Analysis Techniques

Data Analysis	f	%
Descriptive Statistics	44	62.9
Content Analysis	23	32.9
Inferential Statistics	17	24.3
Other Qualitative Analyses	16	22.9
Professional Standards Review	8	11.4
Multiple	30	42.9

Table A-7

Frequencies of Data Sources

Data Source	f	%
Students	53	75.7
Parents	43	61.4
Teachers	43	61.4
Administrators	29	41.4
Governing Body	6	8.6
Counselors	2	2.9
Other	3	4.3
Multiple	53	75.7

Table A-8

Frequencies of Intended Audiences

Intended Audience	f	%
Administrators	53	75.7
Research Community	18	25.7
Governing Body	11	15.7
Teachers	6	8.6
Parents	3	4.3
Counselors	2	2.9
Other	3	4.3
Multiple	18	25.7

Table A-9

Frequencies of Evaluation Concerns

Evaluation Concern	f	%
Curriculum/Instruction	37	52.9
Identification	31	44.3
Organization	31	44.3
General Impressions	30	42.9
Parent/Community Involvement	30	42.9
Outcomes	26	37.1
Staff Development	25	35.7
Adjustment	23	32.9
Resources	19	27.1
Underserved Populations	14	20.0
Foundations	11	15.7
Program Evaluation	11	15.7
Student Evaluation	7	10.0

Table A-10

Frequencies of Reporting Formats

Reporting Format	f	%
General Report	46	65.7
Table	45	64.3
Executive Summary	19	27.1
Other	12	17.1
Multiple	37	52.9

Table A-11

Frequencies of Utility Practices

Utility Practices	f	%
Recommendations	30	42.9
None	21	30.0
Beyond Recommendations	19	27.1

Table A-12

Chi Square Analysis of Evaluation Models by Evaluator Types

		Evaluator Types	
Evaluation Models		Internal	External
Management	f_o	27	10
Centered	f_e	21.5	15.5
Objectives	f_o	11	7
Centered	f_e	10.4	7.6
Other	f_o	2	12
	f_e	8.1	5.9

N = 69
Degrees of freedom = 2
Critical value ($\alpha = .01$) = 9.210
$X^2 = 14.67$

Table A-13

Chi Square Analysis of Data-Gathering Methods by Evaluator Types

		Evaluator Types	
Data-Gathering Methods		Internal	External
Multiple	f_o	20	22
	f_e	24.3	17.6
Questionnaires	f_o	17	3
	f_e	11.6	8.4
Other	f_o	3	4
	f_e	4.1	2.9

N = 69
Degrees of freedom = 2
Critical value ($\alpha = .05$) = 5.991
$X^2 = 8.56$

Table A-14

Chi Square Analysis of Data Sources by Data-Gathering Methods

		Data-Gathering Methodology		
Data Sources		Multiple Methods	Survey	Other
Multiple	f_o	37	14	2
	f_e	32.6	15.1	5.3
Students	f_o	5	2	4
	f_e	6.8	3.1	1.1
Others	f_o	1	4	1
	f_e	3.7	1.7	0.6

N = 70
Degrees of freedom = 4
Critical value ($\alpha = .01$) = 13.277
$X^2 = 16.59$

Table A-15

Chi Square Analysis of Intended Audiences by Evaluation Types

		Evaluation Types		
Intended Audience		Summative	Formative	Other
Administrative	f_o	24	16	5
	f_e	24.4	14.8	5.8
Research	f_o	13	1	1
	f_e	8.1	4.9	1.9
Multiple	f_o	1	6	3
	f_e	5.4	3.3	1.3

N = 70
Degrees of freedom = 4
Critical value ($\alpha = .01$) = 13.277
$X^2 = 14.73$

Table A-16

Chi Square Analysis of Evaluation Models by Intended Audiences

		Intended Audience		
Evaluation Model		Administrative	Research	Multiples
Management	f_o	33	1	3
Centered	f_e	23.8	7.9	5.3
Objective	f_o	8	6	4
Centered	f_e	11.6	3.9	2.6
Other	f_o	4	8	3
	f_e	9.6	3.2	2.1

N = 70
Degrees of freedom = 4
Critical value ($\alpha = .01$) = 13.277
$X^2 = 24.45$

Table A-17

Chi Square Analysis of Intended Audiences by Evaluator Types

		Evaluator Types	
Intended Audience		Internal	External
Administrative	f_o	33	12
	f_e	26.1	18.9
Research	f_o	4	11
	f_e	8.7	6.3
Multiple	f_o	3	6
	f_e	5.2	3.8

N = 69
Degrees of freedom = 2
Critical value ($\alpha = .01$) = 9.210
$X^2 = 12.59$

Table A-18

Chi Square Analysis of Data Analysis Techniques by Intended Audiences

		Intended Audience		
Data Analysis Technique		Administrative	Research	Multiples
Quantitative	f_o	16	12	1
	f_e	18.6	6.2	4.1
Combined	f_o	17	1	6
	f_e	15.4	5.1	3.4
Qualitative	f_o	12	2	3
	f_e	10.9	3.6	2.4

N = 70
Degrees of freedom = 4
Critical value ($\alpha = .01$) = 13.277
$X^2 = 14.56$

Table A-19

Chi Square Analysis of Evaluation Models by Reporting Formats

		Reporting Format			
Evaluation Models		Multiple Formats	Tables	General Reports	Other Formats
Management	f_o	16	12	8	1
Centered	f_e	20.1	6.9	5.8	4.2
Objectives	f_o	12	0	2	4
Centered	f_e	9.8	3.3	2.8	2.1
Other	f_o	10	1	1	3
	f_e	8.11	2.8	2.4	1.7

Degrees of freedom = 6
Critical value ($\alpha = .05$) = 16.812
$X^2 = 17.04$

Table A-20

Chi Square Analysis of Reporting Formats by Evaluator Types

		Evaluator Types	
Reporting Formats		Internal	External
Multiple	f_o	19	18
	f_e	21.5	15.5
Table	f_o	12	1
	f_e	7.5	5.5
General	f_o	7	4
	f_e	6.4	4.6
Other	f_o	2	6
	f_e	4.6	3.4

N= 70
Degrees of freedom = 3
Critical value ($\alpha = .05$) = 7.815
$X^2 = 10.67$

Table A-21

Chi Square Analysis of Data-Gathering Methods by Reporting Formats

Data-Gathering Methodology Models		Reporting Format Multiple Formats	Tables	General Reports	Other Formats
Multiple	f_o	28	2	8	5
Methodology	f_e	23.3	8.0	6.8	4.9
Questionnaire	f_o	6	11	1	2
Centered	f_e	10.9	3.7	3.1	2.3
Other	f_o	4	0	2	1
	f_e	3.8	1.3	1.1	0.8

N = 70
Degrees of freedom = 6
Critical value ($\alpha = .05$) = 16.812
$X^2 = 25.82$

Table A-22

Chi Square Analysis of Data Analysis Techniques by Reporting Formats

Data Analysis Technique		Reporting Format Multiple Formats	Tables	General Reports	Other Formats
Quantitative	f_o	18	6	2	3
	f_e	15.7	5.4	4.6	3.3
Combine	f_o	15	6	2	1
	f_e	13.0	4.5	3.8	2.7
Qualitative	f_o	5	1	7	4
	f_e	9.2	3.2	2.7	1.9

N = 70
Degrees of freedom = 6
Critical value ($\alpha = .05$) = 16.812
$X^2 = 17.24$

Table A-23

Chi Square Analysis of Intended Audiences by Reporting Formats

Data Analysis Techniques		Reporting Format			
		Multiple Formats	Tables	General Reports	Other Formats
Administrative	f_o	22	12	8	3
	f_e	24.4	8.4	7.1	5.1
Research	f_o	11	0	0	4
	f_e	8.1	2.8	2.4	1.7
Multiple	f_o	5	11.9	3	1
	f_e	5.4		1.6	1.1

Degrees of freedom = 6
Critical value ($\alpha = .05$) = 12.592
$X^2 = 13.80$

Table A-24

Chi Square Analysis of Utility Practices by Evaluation Types

Utility Practice		Evaluation Types		
		Summative	Formative	Other
Recommendations	f_o	15	10	5
	f_e	16.3	9.9	3.9
None	f_o	16	2	3
	f_e	11.4	6.9	2.7
Beyond	f_o	7	11	1
Recommendations	f_e	10.3	6.2	2.4

N = 70
Degrees of freedom = 4
Critical value ($\alpha = .01$) = 9.488
$X^2 = 11.$

Table A-25

Chi Square Analysis of Utility Practices by Data-Gathering Methodology

Utility Practice		Data-Gathering Methodology Multiple Methods	 Questionnaire	 Other
Recommendations	f_o	22	5	3
	f_e	18.4	8.6	3
None	f_o	6	13	2
	f_e	12.9	6.0	2.1
Beyond	f_o	15	2	2
Recommendations	f_e	11.7	5.4	1.9

N = 70
Degrees of freedom = 4
Critical value ($\alpha = .01$) = 13.277
$X^2 = 17.15$

Table A-26

Chi Square Analysis of Utility Practices by Reporting Formats

Data Analysis Techniques		Reporting Format Multiple Formats	 Tables	 General Reports	 Other Formats
Recommendations	f_o	20	0	6	4
	f_e	16.3	8.5	4.7	3.4
None	f_o	6	11	1	3
	f_e	11.4	3.9	3.3	2.4
Beyond	f_o	2	2	4	1
Recommendations	f_e	10.3	3.5	3.0	2.2

N = 70
Degrees of freedom = 6
Critical value ($\alpha = .05$) = 16.812
$X^2 = 25.95$

Appendix B

Scale for the Evaluation of Program Evaluation Instruments (*SEPEI*)

SCALE FOR THE EVALUATION OF PROGRAM EVALUATION INSTRUMENTS (SEPEI)

Instrument Name________________________________

File Number ________________________________

Program Goal/Objective Assessed________________________________

____

Instrument Use________________________________

____

Rater________________________ , Date______________

RATING SCALE KEY:

EXCELLENT: The instrument meets all of the criteria standards.

GOOD: The instrument meets most of the described criteria.

FAIR: The instrument meets some of the criteria, or some limited evidence or information is presented.
POOR: The instrument meets none of the criteria, or no supporting evidence is available.
NOT APPLICABLE: Criteria do not apply to the instrument.

DIRECTIONS: The rater first should consult all available sources of information about the instrument included in the data bases of the National Repository of Evaluation Instruments. Then, for each of the evaluation instrument standards included in this rating scale, the rater should check the space corresponding to the appropriate degree (Excellent, Good, Fair, Poor, Not Applicable) to which the instrument meets that standard. Please note that in the criteria standards described below, the term "instrument manual" refers to the formal manual or any directions or other materials that may accompany the instrument. Finally, note that the term "instrument" should be considered in a very broad sense including non-traditional evaluation data-gathering methods.

EVALUATION INSTRUMENT STANDARDS

I. VALIDITY STANDARDS. The rater should consult the instrument manual and published reviews of the instrument to complete this section.

1. Content validity:
The instrument provides a clear definition of the universe represented and provides detailed evidence that the focus of evaluation is measured by the items, including a detailed listing of which items measures which evalution questions.

NA POOR FAIR GOOD EX

2. Construct validity:

a. Experimental construct validity:
Statistical data is presented from a variety of empirical studies using analyses such as factor or correlational analysis, that provide strong support that the instrument accurately reflects the stated underlying construct.

NA POOR FAIR GOOD EX

b. Discriminant construct validity:
Empirical evidence demonstrates that the instrument measures something distinct from what is measured by other instruments of similar format but different constructs. Further, the proposed interpretation of the construct is clearly stated in the manual and distinguished from other interpretations arising from other theories.

NA POOR FAIR GOOD EX

c. Convergent construct validity:
The instrument correlates more closely with other instruments measuring the same construct using different methods than with instruments measuring different constructs.

NA POOR FAIR GOOD EX

3. Criterion validity:

a. Concurrent criterion validity:
Scores on the instrument are related to performance on a separate task or criterion administered concomitantly. The manual also provides user information in terms of the appropriateness of generalizing from the validity information. A given coefficient should be $r = > .70$ for a rating of excellent (Good: $r = .70-.30$; Fair: $r = < .30$; Poor: not reported).

NA POOR FAIR GOOD EX

b. Predictive criterion validity:
Performance on the instrument is related to performance on separate task or criterion administered well after the instrument. A given coefficient should be $r = > .50$ with a criterion relevant interval of at least two months for a rating of excellent (Good: $r = .30-.50$; Fair: $r = < .30$; Poor: not reported).

NA POOR FAIR GOOD EX

BEST COPY AVAILABLE

II. RELIABILITY STANDARDS. The rater should consult the instrument manual and published reviews of the instrument to complete this section.

1. Internal consistency reliability:
NA POOR FAIR GOOD EX

A very high degree of internal consistency is evidenced by an appropriate reliability indicator such as split-half, Kuder-Richardson, or alpha coefficients. This reliability coefficient should be greater than .90 for a rating of excellent (Good: r= .70-.90; Fair: r= <.70; Poor: not reported).

2. Equivalence reliability:
NA POOR FAIR GOOD EX

Alternate form reliability is demonstrated by correlational data between scores on two comparable or parallel forms administered to the same group of respondents at essentially the same time. This reliability coefficient should be greater than .90 for a rating of excellent (Good: r= .70-.90; Fair: r= <.70; Poor: not reported).

3. Test-retest stability reliability:
NA POOR FAIR GOOD EX

Supporting evidence is provided for the stability of scores over a period of time, by correlations on scores obtained by the same groups of respondents on two occasions at least one month apart. For achievement or aptitude instruments, this reliability coefficient should be greater than .90 for a rating of excellent (Good: r= .70-.90; Fair: r= < .70; Poor: not reported). For affective measures, this reliability coefficient should be greater than .70 for a rating of excellent (Good: r= .50-.70; Fair: r=.20-.50; Poor: not reported).

4. Replicability:
NA POOR FAIR GOOD EX

Well defined and controlled procedures and conditions are described such that normative data is well established and the instrument is effectively standardized and useful in order for the user to replicate procedures.

5. Range of coverage:
NA POOR FAIR GOOD EX

A wide range of coverage for the distribution of scores is provided, enabling the raw scores to effectively differentiate among students at the upper end of the scale.

6. Score graduation:
NA POOR FAIR GOOD EX

The raw scores are converted into percentiles, Z-scores, T-scores, grade equivalents, stanines, or other common standard forms of reporting scores for purposes of interpretation.

III. PROPRIETY STANDARDS. The rater should review the instrument and the manual to respond to this section.

1. Ethical/professional:
NA POOR FAIR GOOD EX

The procedures used in administration, instrument content, and recommendations for action conform to accepted ethical assessment practices with due regard to the rights and welfare of those involved in the evaluation.

2. Obligations and disclosure:
NA POOR FAIR GOOD EX

The instrument manual states that parents or guardians of individuals to be assessed must be made aware of all aspects and conditions required by the instrument for administration and assessment and formally agree to these requirements in writing, if appropriate. The manual also encourages an open, direct, and honest presentation of results to examinees, including an explanation of the results.

IV. EXAMINEE APPROPRIATEMESS STANDARDS. The rater should review the actual instrument for this section.

1. Justification/purpose:
NA POOR FAIR GOOD EX

The purpose, intent, or recommended use of the instrument are explained to the respondent in an understandable and straightforward manner in the instructions/directions section.

2. Face validity:
NA POOR FAIR GOOD EX

All items are judged to be appropriate, unambiguous, and suitable for intended respondents in terms of immediate comprehension and degrees of difficulty. Further, the subject matter represented by items is presented in an unbiased fashion that appears relevant and interesting for the respondent.

3. Instructions:
NA POOR FAIR GOOD EX

All instructions are easily understandable and appropriate for the respondent in terms of concepts, syntax, vocabulary, length, and function. The instructions clearly state response form(s) and include sample items that illustrate the necessary skills and item format required for each range of tasks.

4. Format:
NA POOR FAIR GOOD EX

The visual format of the instrument is aesthetically well constructed and easily understandable in terms of its overall organization, layout, print quality, use of illustrations, and consistency of presentation of all ranges of task items.

BEST COPY AVAILABLE

5. Time/pacing: NA POOR FAIR GOOD EX

Adequate and appropriate time limits for responses in relation to the subject matter are provided for the instrument.

6. Recording answers: NA POOR FAIR GOOD EX

The instrument response modes are simple, direct, easily accomplished, and appropriate for the subject matter and the respondent.

V. UTILITY STANDARDS. The rater should review the instrument and manual for this section

1. Audience identification:

The instrument clearly states appropriate type of respondents and evaluation purpose.

Type of respondent_______

Evaluation purpose____

2. Confidentiality:

Provisions for keeping response confidential are provided.

Provision description____

3. Administration:

a. Training: NA POOR FAIR GOOD EX

The instrument can be administered by regular school staff and requires less than one hour of specific training and/or preparation.

b. Group size:

The instrument can be administered to large groups, providing for effective large scale evaluation.

____ Large group (can be given to 10 or more individuals)
____ Small group (limited to 2-10 individuals)
____ Individual administration necessary

c. Time:

The instrument requires an appropriate amount of time for ease of administration and is organized to allow for appropriate breaks given type of respondent.

Total time required:

____ < 20 minutes ____ 20-40 minutes ____1-2 hours ____>2 hours

d. Manual quality: NA POOR FAIR GOOD EX

The instrument manual is legible, well-organized, consistent, easy to use, and thorough in its directions for administration. The manual also provides comprehensive information in regard to the purpose and uses of the instrument and reliability and validity data.

4. Scoring:

a. Scoring ease:

The instrument can be easily and objectively scored on-site

____ Must be hand-scored
____ May be hand-scored or machine-scored
____ Must be machine-scored
____ Must be sent to publisher for scoring.

b. Score conversion: NA POOR FAIR GOOD EX

A simple, one-step process of accurately converting raw scores to normed/interpreted scores is provided by use of clearly explained tables, graphs, or scoring keys.

c. Report clarity and distribution: NA POOR FAIR GOOD EX

Examinee scores or performance are presented in clear and self-explanatory terms free of unnecessary jargon and are distributed to respondents and others in a timely fashion.

BEST COPY AVAILABLE

5. Interpretation:

a. **Evaluator training:**

The instrument can be directly and immediately interpreted by regular school personnel relative to a specified norm group or standard. The necessary qualifications for persons interpreting results and guidelines are explicitly stated in the test manual.

____ Teachers can easily interpret results
____ Special trained individuals needed for interpretation.

b. **Norm range:** NA POOR FAIR GOOD EX

Data is provided for purposes of interpretation indicating that the instrument has been normed on a broad range of educational ability or is applicable to groups falling at the upper end of the continuum.

c. **Norm timing:**

The norms for the instrument are current (within last five years). Date of norming:____________________

d. **Norm groups:**

Adequate norms are provided for special populations of students.

____ Norms are comprehensive but no specific norms on gifted populations
____ Norms are provided for special populations:
    ____ handicapped
    ____ economically disadvantaged
    ____ minority: ____ black ____ Native American ____ Asian ____ Hispanic

6. **Evaluation:** NA POOR FAIR GOOD EX

Explicit guidelines are described for using results to make objective and valid assessments of student and/or program performance.

7. **Cost effectiveness:** NA POOR FAIR GOOD EX

Supportive evidence exists that justifies the resources expended for the use of the instrument in terms of information gained.

8. **Political viability:** NA POOR FAIR GOOD EX

The instrument is considered as an accepted means of evaluating gifted programs by various interest groups and professional education/measurement/evaluation associations.

VI. GENERAL RATER COMMENTS: (Please use the reverse side if necessary)

BEST COPY AVAILABLE

Appendix C

Technical Manual for the *Scale for the Evaluation of Program Evaluation Instruments* (*SEPEI*)

Technical Manual for the *Scale for the Evaluation of Program Evaluation Instruments* (*SEPEI*)

Prepared by
A. Christian Lundberg
Scott L. Hunsaker
Carolyn M. Callahan

Published by the
The National Research Center on the Gifted and Talented
University of Virginia Site

Dr. Carolyn M. Callahan
Site Director

Technical Manual for the *Scale for the Evaluation of Program Evaluation Instruments* (*SEPEI*)

I. Introduction

Given the diversity in types of programs for the gifted, the wide variety of goals/outcome statements, and the resulting confused state of the art concerning the reliability and validity of instruments used for the evaluation of gifted programs, it is no wonder that local educational administrators and teachers are perplexed when faced with the prospect of making informed choices about program evaluation instruments. The most common problem concerns the reliability, validity, and utility instruments which might be used at the local school district level.

There has been little done to provide comprehensive reviews and assessments of instruments for the specialized purpose of evaluating gifted programs. Although there have been articles dealing with evaluation of gifted programs (Callahan, 1983; Carter, 1986), there still is little information available, other than that found in general test reviews, concerning the reliability and validity instruments used to evaluate the process and outcomes of programs designed for gifted students. Instruments which are not published or are locally developed are most often not included in any "collections" that may be available to local schools. The few existing collections do not generally include non-traditional means of assessment such as portfolio reviews, peer rating, or evaluations of student products. In response to this pervasive need, a major part of the mission of The National Research Center on the Gifted and Talented (NRC/GT) was specifically devoted toward the collection of evaluation instruments and the development of a rating scale that would assess existing gifted program evaluation instruments for the variety of situations in which they might be used.

The *Scale for the Evaluation of Program Evaluation Instruments* (*SEPEI*) was designed by project staff at the University of Virginia site of the NRC/GT with the intent to provide a comprehensive review of the effectiveness, appropriateness, and overall value of all currently available instruments and procedures used for the purpose of evaluating gifted programs. These ratings of instruments for specific uses based on program evaluation needs were assembled into a National Repository of Instruments that serves as a resource for local school districts desiring information concerning the reliability, validity, utility, and appropriateness of an instrument.

II. Uses of *SEPEI*

Gallagher (1988) has included program evaluation among the priorities he identifies as crucial for the continued improvement of gifted education. Determining the merits of various instruments that will be part of a comprehensive program evaluation is needed prior to the conduct of the evaluation. An evaluation which draws conclusions or makes recommendations based on data from unreliable or invalid instruments is a dangerous procedure.

Individuals and local school systems interested in evaluating their program with sound instruments can contact the National Repository of Instruments of the NRC/GT for information and advice as to the reliability and validity of instruments and procedures through the comprehensive *SEPEI* ratings conducted by the research staff of the NRC/GT. A wide range of evaluation instruments have been evaluated by NRC/GT staff. However, any repository is limited by the submissions of cooperating groups, and more importantly, every assessment tool should be carefully considered for reliability and validity for the particular situation and decisions which will be made using the instrument. Hence, educators may not always find information on a particular assessment tool or the uses in a particular circumstance may be new and unique. Hence, educators may wish to use the SEPEI for purposes of either evaluating locally developed instruments, a situation-specific use of an instrument, or as a guide in the development of any new instruments.

III. Overview of Instrument Development

Content Validity of *SEPEI*

A review of the literature was conducted to determine the most important standards or criteria that should be met by gifted program evaluation instruments. The main sources consulted included *Guidelines for Test Use* (Brown, 1980), *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, National Council on Measurement in Education, 1985), *Standards for Evaluations of Educational Programs, Projects, and Materials* (Joint Committee on Standards for Educational Evaluation, 1981), and *Principles of Educational and Psychological Measurement and Evaluation* (Sax, 1989). The instrument was eventually based on models of instrument evaluation forms from the Evaluation Technologies Program of the Center for the Study of Education and the Humanizing Learning Program of Research for Better Schools, Inc. (Hoepfner, Strickland, Jansen, & Patalino, 1970), which have demonstrated promise in providing a full and understandable assessment of the reliability and validity of an instrument.

From this review of the literature, a comprehensive instrument was constructed by project staff of the NRC/GT. Items, or what are termed "criteria standards," were developed for five major areas of assessment: 1) Validity Standards, 2) Reliability Standards, 3) Propriety Standards, 4) Respondent Appropriateness Standards and 5) Utility Standards. These standards are amplified in the descriptions presented below:

Identification Instrument Standards of *SEPEI*

1. **Validity Standards.** These standards are concerned with the presupporting question that underlies all other aspects of instrument validity: "How well does the instrument measure, for its intended respondent and purpose, the specific construct that it claims to represent?" Standards for assessment included here are content, construct, and criterion validity.
2. **Reliability Standards.** Ratings for these standards are concerned with the extent to which the instrument is consistent and accurate in its operation and in providing information for any particular occasion that it is used. Internal consistency, equivalence, stability, and replicability are examples of criteria standards included in this section.
3. **Propriety Standards.** The degree to which an instrument openly addresses fundamental ethical and professional considerations of testing, measurement, and evaluation is perhaps the most important indicator of the worthiness of an instrument. These standards, which also include obligations and disclosure must be met by any instrument that is used for the purpose of psychological testing or program evaluation.
4. **Respondent Appropriateness Standards.** Ratings in this category are concerned with the suitability of an instrument for the individual or group that will either be assessed or will be involved in the completion of that instrument. Standards under this heading include the appropriateness of

instruction, face validity, method of recording answers, format time/pacing, and justification/purpose.

5. **Utility Standards.** These standards are concerned with the more practical considerations involved in administering and using a test or other assessment tool, including scope and time of administration, administrator training, manual quality, scoring procedures, guidelines for interpretation and decision making (including norming information), and political viability (the instrument's "acceptability" among professionals and interest groups).

Each criterion standard or item for these major categories was written in the form of a paradigm or "best case scenario," with each standard to be rated by the degree to which the instrument met that standard: "Excellent," "Good," "Fair," "Poor," or "Not Applicable." The possible rating responses are further described below:

RATING SCALE KEY
Excellent: The instrument meets all of the criteria standards.
Good: The instrument meets most of the described criteria standards.
Fair: The instrument meets some of the criteria standards or some limited evidence or information is presented.
Poor: The instrument meets none of the criteria or no supporting evidence is available.
Not Applicable: The criteria do not apply to the instrument.

As the SEPEI criteria standards are relatively complex, where appropriate, additional guidelines and measurement rules of thumb were included in the criteria descriptions to aid raters in making more accurate judgements. In addition a final section of the scale was provided for "General Rater Comments" to allow raters to include a brief summary of their overall impressions and recommendations concerning the instrument. It is hoped that any instrument will conform to all of the statements described in the scale. However, because of the difficulty involved in designing an instrument to provide a full and clear picture for all kinds of evaluation instruments (including non-standardized measures such as opinion surveys), all kinds of respondents (e.g., student, teacher, parent), and for various program types, the response choice of "Not Applicable" was included if a particular standard may not apply to a particular instrument.

To further determine the content validity of SEPEI, the instrument was submitted for formative evaluation on two occasions to a seven member panel of individuals in the fields of education of the gifted, special populations of students, and psychometrics from the University of Virginia with expertise in measurement and evaluation. Each of these individuals was asked to carefully assess the content of the instrument for its comprehensiveness (including duplications and omissions), clarity, and utility and relevancy for its intended purpose. Suggestions received by these reviewers on each occasion were assessed and appropriate recommendations for revisions were incorporated into the final version of SEPEI.

Reliability of the *SEPEI*

Studies to establish inter-rater reliability were conducted on two instruments during the spring of 1991. A panel of four raters participated, graduate students in educational psychology and two faculty members with experience in tests and measurements and evaluating gifted programs. These studies were conducted by having each rater independently rate a test which had been submitted to the pool of available instruments. The inter-rater reliability was assessed as the percentage agreement (PA) for 1) the highest agreement on any one response choice for each item on the rating scale (Actual PA) and 2) the highest agreement on any two adjoining response choices for each item on the rating scale (PA Within Two). For example, an item might have 75% of the raters rating a test as good on an item, but the other rater rated it fair on the same item. The Actual PA would be 75%, the PA Within Two would be 100%. If 50% had rated it fair, 25% rate it good and 25% poor, the Actual PA would be 50% and the PA within Two would be 75%. The two instruments assessed were *Ross Test of Higher Cognitive Processes* and the *Cornell Critical Thinking Appraisal.*

In each rating trial, raters were given the instrument to be assessed and also published test reviews, and all available recent research pertaining to the reliability and validity of the instrument for use in conducting the assessment of the test. The results of the raters are presented in Appendix D of this document.

IV. Directions for Using the *Scale for Evaluation of Program Evaluation Instruments*

General Instructions

Before completing the scale, the rater first should consult all available sources of reliability/validity information and other reviews of the instrument. The rater should also collect any pertinent information relating to reliability, validity, and program information if the instrument is being reviewed in the context of a local gifted program. Then, for each of the identification instrument standards included in this rating scale, the rater should check the space corresponding to the appropriate degree ("Excellent, Good, Fair, Poor, Not Applicable") to which the instrument meets that standard (SPECIAL NOTE: "Not Applicable" should only be used for rare instances when a standard may not apply due to the nature of the instrument.) Please note that in the criteria standards described on the scale, the term "instrument manual" refers to the formal manual or any directions or other materials that may accompany the instrument. Finally, note that the term "instrument" always should be considered in very broad sense, thereby including non-standardized practices such as auditions, portfolios, performance rating scales, and questionnaires.

At the local level, it is recommended that several individuals complete the scale in order to obtain a larger base of information for a more thorough assessment of the instrument in regard to its particular use. It is important to remember that the *Scale for the Evaluation of Program Evaluation Instruments* is not designed to issue an overall "score" for the instrument being rated. Rather, it is designed to provide a complete "report" and critical evaluation of an instrument to promote a fuller understanding of the merits and shortcomings of the instrument in light of its use for purposes of evaluating gifted programs.

Supplementary Instructions

1. Always make sure that you first review the instrument before completing the rating scale in order to gain a sense of the instrument's "face" validity, propriety, utility, and appropriateness.
2. Please note that "NA" should only be used for "not applicable" (e.g., the criterion does not apply to the instrument). Sometimes a criterion may not apply to an instrument (e.g., parallel form are not furnished by the instrument, hence **equivalence reliability** (**II.2** receives a "NA") but in most cases all of the criteria in the scale should be addressed by the instrument rated. If desired information for a criterion is not given by the instrument, then "POOR" should be checked.
3. When completing the **Ethical/Professional** standards criterion (**III.1**) raters should approach the item by thinking, "What does the instrument say that it is going to do,

and how well does it inform the reader as to how it will openly and accurately carry out its claims?"

4. Please note for the **Respondent Appropriateness Standards (IV)** that the **Justification/Purpose, Instructions, Format, and Time/Pacing** standards (**IV. 1, 3, 4, 5**) and criteria all involve "judgement call" responses, and may represent a source of rater bias in the scale. It is therefore very important to keep in mind the instrument's intended respondent when completing these items in order to provide the most accurate assessments. All raters should consider the extent to which the instrument "matches" with the respondents for such items.
5. A source of bias inherent in the **Utility Standards** section (**V**) of the SEPEI is the pronounced emphasis on the **efficiency** of the use of an instrument. For example, throughout the construction of this section, items were designed with the assumption that the local gifted teacher is the most efficient (if not always effective) individual to perform the administration (**Utility Standard V.3.a**). Further, in terms of group size (**Utility Standard V.2.b**) and length of time required to use the instrument (**Utility Standard V.2.c**), it is assumed that large group evaluation and minimal time of instrument administration are appropriate standards for the highest rating responses. Extended direction for performing ratings on items such as these are provided in the criterion standards of the instrument.
6. When answering the **Reliability Standards (II)** and **Validity Standards (I)** sections of the scale, the rater should remember the purpose and recommended use of the instrument as well as the nature of the instrument itself. What the instrument claims to be and to do has a direct influence upon how the authors attempt to establish its credibility. For example, if the test is intended for use as a predictive instrument, then there should be some evidence of **predictive criterion validity** (**I.3.b**). And, if the test claims to be different than other tests, it should substantiate this by evidence for **discriminant construct validity** (**I.2.b**). (**Convergent construct validity** (**I.2.c**) is seen when the instrument intends to measure the same domain or construct as other tests, but does so by a different method). Please also be aware that instrument developers alternately use a discriminant or a convergent approach to prove their points. Always check what criterion are used by the authors to establish the instrument's validity and how the authors are comparing their instruments to the criterion.
7. Again, the rater should consider the intended respondent audience when answering **Utility Standards for Audience identification, Group size, and Time** (**V.1, 2.b, & 2.c**). These data should be clearly stated in the instrument manual.

References

American Educational Research Association. (1985). *Standards for educational and psychological testing.* Washington, DC: American Psychological Association, National Council on Measurement in Education.

Callahan, C. M. (1983). Issues in evaluating programs for the gifted. *Gifted Child Quarterly, 27,* 3-7.

Brown, F. G. (1980). *Guidelines for test use: A commentary on the standards for educational and psychological testing.* Washington, DC: National Council on Measurement in Education.

Carter, K. R. (1986). A cognitive outcomes study to evaluate curriculum for the gifted. *Journal for the Education of the Gifted, 10,* 41-55.

Gallagher, J. J. (1988). National agenda for educating gifted students: Statement of priorities. *Exceptional Children, 55,* 107-114.

Hoepfner, R., Strickland, G., Jansen, P., & Patalino, M. (Eds.). (1970). *CSE elementary school test evaluation.* Los Angeles: UCLA Graduate School of Education, Center for the Study of Evaluation.

Joint Committee on Standards for Educational Evaluation. (1981). *Standards for evaluations of educational programs, projects, and materials.* New York: McGraw-Hill.

Sax, G. (1989). *Principles of educational and psychological measurement and evaluation.* Belmont, CA: Wadsworth Publishing.

Appendix D

SEPEI Inter-rater Item Descriptive Statistics for the *Cornell Test of Critical Thinking* and the *Ross Test of Higher Cognitive Processes*

SEPEI Inter-rater Item Descriptive Statistics for the *Cornell Test of Critical Thinking*

Standard Rated	N of Raters	Mean Rating	*SD*	Minimum Rating	Maximum Rating
Validity Standards:					
content	4	3.00	0.82	2	4
construct-experimental	4	2.00	0.82	1	3
discriminant	4	1.00	0.82	0	2
convergent	4	1.75	1.26	0	3
criterion-concurrent	4	1.00	1.41	0	3
predictive	3	1.00	1.00	0	2
Reliability Standards:					
internal consistency	4	2.75	0.50	2	3
equivalence	4	0.75	0.96	0	2
stability	4	1.25	1.258	0	3
replicability	4	2.50	1.000	2	4
range of coverage	4	2.00	0.000	2	2
score graduation	4	2.25	0.500	2	3
Propriety Standards:					
ethical/professional	4	2.50	0.577	2	3
obligations/disclosure	4	1.50	0.577	1	2
Examinee/Appropriateness Standards:					
justification/purpose					
face validity	2	3.00	0.00	3	3
instructions	1	4.00	0.00	4	4
format	3	3.00	0.00	3	3
time/pacing	2	2.00	0.00	2	2
recording answers	2	3.00	0.00	3	3
	3	3.67	0.58	3	4
Utility Standards:					
administration-training	4	3.75	0.50	3	4
manual quality	4	2.50	0.58	2	3
score conversion	4	1.75	0.50	1	2
report clarity/distribution	4	2.00	1.16	1	3
norm range	4	1.75	1.26	0	3
evaluation	3	0.67	0.58	0	1
cost effectiveness	4	2.75	1.26	1	4
political viability	4	0.50	1.00	0	2

For further information on the reliability of the *SEPEI*, consult Part I.

BEST COPY AVAILABLE

SEPEI Inter-rater Item Descriptive Statistics for the *Ross Test of Higher Cognitive Processes*

Standard Rated	N of Raters	Mean Rating	*SD*	Minimum Rating	Maximum Rating
Validity Standards:					
content	4	2.50	1.29	1	4
construct-experimental	4	1.75	0.95	1	3
discriminant	4	2.75	0.50	2	3
convergent	4	1.00	0.00	1	1
criterion-concurrent	4	0.75	0.50	0	1
predictive	4	4.00	0.00	4	4
Reliability Standards:					
internal consistency	4	0.00	0.00	0	0
equivalence	4	2.00	0.28	1	3
stability	4	2.50	1.00	2	4
replicability	4	2.50	1.29	1	4
range of coverage	4	3.25	0.50	3	4
score graduation	4	3.00	0.82	2	4
Propriety Standards:					
ethical/professional	4	1.75	0.96	1	3
obligations/disclosure	4	1.50	0.00	1	3
Examinee/Appropriateness Standards:					
justification/purpose					
face validity	4	3.00	0.00	3	3
instructions	4	3.25	0.96	2	4
format	4	3.50	0.58	3	4
time/pacing	4	2.50	0.58	2	3
recording answers	4	3.25	0.50	3	4
	4	2.50	1.00	1	3
Utility Standards:					
administration-training	4	3.00	0.00	3	3
manual quality	4	3.50	0.58	3	4
score conversion	3	1.33	1.53	0	3
report clarity/distribution	4	2.50	1.73	0	4
norm range	4	0.75	0.50	0	1
evaluation	4	1.50	1.00	0	2
cost effectiveness	4	0.50	0.58	0	1
political viability	4	1.00	1.16	0	2

Kendall Coefficient of Concordance for the *Ross Test of Higher Cognitive Processes*

Variable	Mean Rank
Validity Standards:	
content	16.33
construct-experimental	14.17
discriminant	18.33
convergent	8.17
criterion-concurrent	6.67
predictive	27.83
Reliability Standards:	
internal consistency	2.67
equivalence	12.17
stability	18.33
test-retest stability	14.33
replicability	21.50
range of coverage	20.67
score graduation	
Propriety Standards:	
ethical/professional	10.00
obligations/disclosure	13.17
Examinee/Appropriateness Standards:	
justification/purpose	21.50
face validity	21.83
instructions	23.83
format	16.50
time/pacing	23.83
recording answers	17.17
Utility Standards:	
administration-training	21.50
manual quality	23.83
score conversion	11.00
report clarity/distribution	15.00
norm range	6.67
evaluation	10.83
cost effectiveness	4.50
political viability	6.00

Cases	W	Chi-square	D.F.	Significance
3	.6964	58.4960	28	.0006

Appendix E

A Typical Response to Request for Data From the Evaluation Database

NRC
G/T

THE NATIONAL RESEARCH CENTER
ON THE
GIFTED AND TALENTED

Joseph S. Renzulli
Director
University of Connecticut

Carolyn M. Callahan
Associate Director

Curry School of Education
University of Virginia
405 Emmet Street
Charlottesville, VA 22903
TEL (804) 982-2849
FAX (804) 924-0747

Francis X. Archambault
Associate Director

1785
The University of Georgia

Mary M. Frasier
Associate Director

Robert J. Sternberg
Associate Director

September 20, 1994

Darrell Cravitz
Director of GT Programs
2908 Stradford Lane
Blacksburg, Va. 24060

Dear Mr. Cravitz:

Thank you for your interest in the National Repository of Identification and Evaluation Instruments. Based on our earlier correspondence, I am sending you copies of the EVALDES (evaluation design database reports) and EVALUTIL (matching instruments with evaluation questions). Hopefully, these will be of some use to you. If you have any further questions, do not hesitate to contact me at the main center phone number.

Sincerely,

Johann H. Lee
Database Manager

Funded by the Office of Educational Research and Improvement, United States Department of Education

BEST COPY AVAILABLE

Figure 2. List of Standardized Instruments Used but Unrelated to a Specific Evaluation Question

Animal Crackers
Career Decision Making Skills
California Achievement Test
Children's Task Persistence
Kaufman Assessment Battery for Children
Kit of Factor Referenced Cognitive Tests
Piers Harris Children's Self Concept Scale
Preliminary Scholastic Aptitude Test
Role Category Test
Ross Test of Higher Cognitive Processes
Scholastic Aptitude Test
Self-Concept and Motivation Inventory
SRA Achievement Test
TAAS Criterion-Referenced Test
Thinking Creatively in Action and Movement
Torrance Tests of Creative Thinking
Williams Test of Divergent Thinking
unspecified achievement tests

BEST COPY AVAILABLE

Figure 3. Standardized Instruments Used to Assess Program Evaluation Questions.

Category of Evaluation Question	**Name of Instrument**	#
achievement	California Achievement Test	2
	Clymer Barrett	1
	Comprehensive Test of Basic Skills	5
	Iowa Tests of Basic Skills	1
	Metropolitan Achievement Test	1
	Peabody Picture Vocabulary Test	1
	Preliminary Scholastic Aptitude Test	1
	Scholastic Aptitude Test	1
	Sequential Tests of Educational Progress, Series III	1
	Stanford Achievement Test	3
	Texas Educational Assessment of Minimal Skills	1
	Test of Academic Aptitude	1
aptitude	Developing Cognitive Abilities Test	1
attitudes toward others	School Situation Survey	1
autonomy/responsibility	Intellectual Achievement Responsibility Scale	1
creativity	Something About Myself	1

	Student Product Assessment Form	1
	Test of Creative Potential	1
	Thinking Creatively in Action and Movement	1
	Torrance Test of Creative Thinking	2
	Torrance Test of Creative Thinking - Demonstrator Form	1
	Wallach-Kogan Creativity Instrument	2
general academic outcomes	California Achievement Test	1
	Ross Test of Higher Cognitive Processes	1
	Scholastic Aptitude Test	1
	Survey of School Attitudes	1
	Test of Cognitive Skills	1
general affective outcomes	Dimensions of Self Concept Inventory	1
general program outcomes	California Achievement Test	2
	Clymer Barrett	1
	Comprehensive Test of Basic Skills	1
	Criterion Referenced Talent Tests	1
	Preliminary Scholastic Aptitude Test	1
	Stanford Achievement Test	1
	Torrance Test of Creative Thinking	1
general student growth	California Achievement Test	1
	Cornell Critical Thinking	1
	Ross Test of Higher Cognitive Processes	2

	Torrance Tests of Creative Thinking	1
identification	California Achievement Test	3
	Cognitive Abilities Test	1
	Comprehensive Test of Basic Skills	1
	Culture Free SEI	1
	Iowa Tests of Basic Skills	2
	Matrix Analogies Test	1
	Otis-Lennon Mental Abilities Test	1
	Otis-Lennon School Abilities Test	1
	Scales for Rating the Behavioranl Characteristics of Superior Students	1
	Scholastic Aptitude Test	1
	Stanford Achievement Test	1
	Stanford-Binet	1
	Structure of Intellect Gifted Screening Form	1
	TDT	1
	Test of Cognitive Skills	1
	Wechsler Intelligence Scale for Children-Revised	2
locus of control	James' Internal/External Locus of Control Scale	1
research skills	GAIN Teacher Assessment of Student Research Skills	1
self-concept	Coopersmith Test of Self-Esteem	1
	Harter Self-Perception Profile	1
	ME Scale	3

	Piers Harris Children's Self Concept Scale	1
	Revised Janis-Field Feelings of Inadequacy Scale	1
	Self-perception Inventory	1
student perceptions	Quality of School Life	2
study habits	Survey of Study Habits and Attitudes	1
thinking skills	Cognitive Abilities Test	1
	Criterion Referenced Talent Tests	2
	Developing Cognitive Abilities Test	2
	Ross Test of Higher Cognitive Abilities	8
	Sequential Tests of Educational Progress, Series III	1
	Stanford-Binet	1
	Talent Assessment Checklist	1
	Texas Educational Assessment of Minimal Skills	1
	Watson-Glaser Critical Thinking Appraisal	2
	Wechsler Intelligence Scale for Children-Revised	1

File Number	BOR-NRC-044
Bib. Entry	Borich, G. D. (1980). A state of the art assessment of educational evaluation. Austin, TX: University of Texas. (ERIC Document Reproduction Service No. ED187 717).
Confidentiality	
Eval. Design	The history of evaluation is discussed in terms of the effects of the behavioral objectives, movement, logic of physics, curriculum reform movement, Elementary and Secondary Education Act, accountability movement highlighted. Several definitions of evaluation are presented: evaluation as measurement, determining consequence, professional judgment, and applied research. Evaluation models are highlighted: the Discrepancy Model, the State Model, and CIPP with a comparison table featured. Emerging trends included are: decision-oriented evaluation, value-oriented, naturalistic (responsive, judicial, transactional, connoisseurship, illumination), and a system-oriented approach. Implications are discussed.
Eval. Type	
Eval. Model	
Evaluator Type	
Data Gen./Analysis	
Data Gath. Methods	
Data Analysis-Tech.	
Utility Info.	
Intended Audiences	
Reporting Format	
Utility Info. Avail	
Cross-References	
Comments	

File Number	CAR-NRC-071
Bib. Entry	Carr, C., Castilhos, M., Davis, D., Synder, M., & Stecher, B. (1982). Evaluation Studies: Cost-benefit analysis in educational evaluation. Studies in Educational Evaluation, 8, 75-85.
Confidentiality	
Eval. Design	Cost-benefit analysis as applied to evaluation is discussed for a specific evaluation of a graduate school of education. Costs and benefits were first identified by estimation of direct and indirect costs. Benefits were weighted and a cost-benefit ratio was established. Practical concerns are identified.
Eval. Type	
Eval. Model	
Evaluator Type	

BEST COPY AVAILABLE

Data Gen./Analysis	
Data Gath. Methods	
Data Analysis Tech.	
Utility Info.	
Intended Audiences	
Reporting Format	
Utility Info. Avail	
Cross-References	
Comments	

File Number	CAR-NRC-118
Bib. Entry	Carter, K. R. (1991). *A model for evaluating programs for the gifted under non-experimental conditions*. Unpublished manuscript, University of Northern Colorado, Greeley.
Confidentiality	
Eval. Design	Model components include "ex post facto design with intact groups, comparative evaluation, strength of treatment and multiple outcome assessment from flexible data sources."
Eval. Type	SM
Eval. Model	PC
Evaluator Type	EX
Data Gen./Analysis	
Data Gath. Methods	
Data Analysis Tech.	
Utility Info.	Assumptions underlying this model include: meaningful data can be obtained wihtout tightly controlled experimental conditions, instrumentation is already available or can be constructed to measure outcomes, and comparison groups can be obtained as a test of curricula for the gifted. A detailed example of how the model can be used is presented.
Intended Audiences	AD
Reporting Format	
Utility Info. Avail	

BEST COPY AVAILABLE

Cross-References	Callahan, C. M. (1983). Issues in evaluating programs for the gifted. *Gifted Child Quarterly*, *27*(1), 3-7. Callahan, C. M. (1986). Asking the right questions: The central issue in evaluating programs for the gifted and talented. *Gifted Child Quarterly*, *30*(1), 38-42. Carter, K. R. (1986a). Evaluation design: Issues confronting evaluators of gifted programs. *Gifted Child Quarterly*, *30*(2), 88-92. Carter, K. R. (1986b). Measuring program outcomes: Suggestions to evaluators of gifted programs. *Illinois Council for the Gifted Journal*, *5*, 38-40. Carter, K. R. (1986c). A cognitive outcomes study to evaluate curriculum for the gifted. *Journal for the Education of the Gifted*, *10*(1), 41-55. Carter, K. R., & Swanson, L. (1990). An analysis of the most frequently cited journals since the Marland report: Implications for researchers. *Gifted Child Quarterly*, *34*(3), 116-123. Carter, K. R., & Hamilton, W. (1985). Formative evaluation of gifted programs: A process and model. *Gifted Child Quarterly*, *29*(1), 5-11.
Comments	
File Number	CLI-NRC-123
Bib. Entry	Clinkenbeard, P. R. (1992, April). *Using qualitative methods to evaluate programs for the gifted.* Paper presented at the annual convention of the American Educational Research Association, San Francisco, CA.
Confidentiality	

Eval. Design	Qualitative methods are described and applied to

BEST COPY AVAILABLE

evaluation. Qualitative and qualitative methodology are compared in the call for using qualitative method for evaluation. This reasoning is applied to using qualitative methodology for gifted program evaluation in view of problems and solutions in evaluating programs for the gifted. Problems in evaluating programs for the gifted include: 1) instruments have psychometric problems inherent in instrument or inapplicability to gifted population (Callahan, 1992); 2) lack of ready-made, valid instruments to measure gifted program goals; 3) low ceilings, lack of gifted norms, regression to mean effect, unreliability of gain scores, difficulties in using true experimental design (Tannebaum, 1983; Borland, 1989). Other authors (VanTassel-Baska, 1989; Borland, 1989; Renzulli, 1975) address issues by focusing on perspective of evaluators. Qualitative methodology, on the other hand, focuses on the perspective of the participants. Qualitative methodology also avoids the psychmetric and design problems previously discussed. However, qualitative methodology is especially suited to gifted program evaluation because 1) methods are more appropriate for program geared toward independent study and individualized student outcomes; 2) can better illustrate results of complex goals; 3) will reveal unanticipated program results; and 4) can determine if programs are "qualitatively different." Examples of the rest of qualitative methods in gifted education are provided.

Eval. Type	
Eval. Model	
Evaluator Type	
Data Gen./Analysis	
Data Gath. Methods	QU; IN; OB; DA
Data Analysis Tech.	CA; OL
Utility Info.	
Intended Audiences	
Reporting Format	
Utility Info. Avail	UT
Cross-References	
Comments	

File Number	COO-NRC-081
Bib. Entry	Cooley, W. W., & Lohnes, P. R. (1977). Value and outcome attributions in educational evaluation. *Education and Urban Society*, *9*, 493-507.
Confidentiality	

Eval. Design	Two problems in evaluation are identified: 1) attributing value to outcome measures; and 2) attributing outcome effects to particular school

BEST COPY AVAILABLE

practices. For problem #1, the author contend that research into the "transfer value of outcomes" needs to be conducted. For problem #2, they identified 4 measures related to instructional or treatment variable: opportunity, motivator,, structure, and instructional event measures. Thus, performance of students may be based on something other than the treatment. Research is needed on developing a model of classroom environments.

Eval. Type
Eval. Model
Evaluator Type
Data Gen./Analysis
Data Gath. Methods
Data Analysis Tech.
Utility Info.
Intended Audiences
Reporting Format
Utility Info. Avail
Cross-References
Comments

File Number | JOH-NRC-022
Bib. Entry | Johnson, R. T. & Thomas, W. P. (1979). User experiences in implementing RMC Title I evaluation models. (ERIC Document Reproduction Service No. ED178 612).

Confidentiality
Eval. Design | The implementation of the RMC evaluation models by state and local education agencies is described. These models varied widely according to selection, adminstration, and scoring of tests, and data analysis and aggregation. Problems in model implementation were either procedural, clerical, or analytical. Suggestions for improvement are outlined for each problem area.

Eval. Type
Eval. Model
Evaluator Type
Data Gen./Analysis
Data Gath. Methods
Data Analysis Tech.
Utility Info.
Intended Audiences
Reporting Format

Utility Info. Avail
Cross-References
Comments

File Number | MAH-NRC-058

BEST COPY AVAILABLE

Bib. Entry	Maher, C.A., & Mossip, C. E. (1984, May). An evaluation system for development and improvement of educational programs for gifted children in the public schools. Educational Technology, 39-44.
Confidentiality	
Eval. Design	The Program Analysis and Review System (PARS) is described. This system encourages a wide range of evaluation information to be collected for improved decision-making, empirically as applied to gifted programs. PARS meets for needs of evaluation for those programs including its emphasis on collaboration between the evaluator and the manager; it requires program parameter specification; its form is on the process; and it uses multiple measures and perspectives to determine program outcome. PARS consists of 3 steps: program specification (client, client needs, program goals, indication of goal attainment, resource components of the program, assumptions linking components to the goals, and the evaluation design; program documentation; and program outcome determination.
Eval. Type	
Eval. Model	
Evaluator Type	
Data Gen./Analysis	
Data Gath. Methods	
Data Analysis Tech.	
Utility Info.	
Intended Audiences	
Reporting Format	
Utility Info. Avail	
Cross-References	
Comments	

File Number	NIE-NRC-018
Bib. Entry	Nielson, L., & Turner, S. D. (1983). Program evaluation as an evolutionary process. Evaluation Review, 7, 397-405.
Confidentiality	
Eval. Design	Evaluation must change as programs change. Implication is that both the evaluation questions and designs will change, thus different evaluation approaches will be utilized. Two examples are used as illustration.
Eval. Type	
Eval. Model	
Evaluator Type	
Data Gen./Analysis	
Data Gath. Methods	
Data Analysis Tech.	
Utility Info.	

BEST COPY AVAILABLE

Intended Audiences Reporting Format Utility Info. Avail Cross-References Comments	
File Number	NOR-NRC-032
Bib. Entry	Norris, S. P. (1986). Evaluating critical thinking ability. History and Social Science Teacher, 21, 135-146.
Confidentiality	
Eval. Design	Implications for evaluating critical thinking ability include: 1) evaluation must be on process, not product; 2) a clear conception of "critical-thinking abilities" must be established; 3) it must not be assumed that critical thinking ability can be transferred across situations; and 4) student evaluation of critical thinking must be examined. Several issues are discussed in designing evaluations of critical thinking ability: collecting information via individual/group tests, or essay/objective tests; nonstandard uses of tests; naturalistic observations; quality and meaning of collected information; reliability and validity, and quality of teacher-made and commercial tests. Several tests highlighted included: Watser-Glaser Critical Thinking Appraisal, New Jersey Test of Reasoning Skills, Test on Appraising Observations, and The Ennis-Weir Critical Thinking Essay Test.
Eval. Type	FO; SM
Eval. Model Evaluator Type Data Gen./Analysis Data Gath. Methods Data Analysis Tech.	
Utility Info.	Uses of this evaluation information can include: decisions concerning instruction, decisions concerning teacher-made or program effectiveness, and decisions regarding shaping of programs and staff development. To make decisions about program effectiveness, however, a comparison group similar to experimental group is necessary.
Intended Audiences Reporting Format	

BEST COPY AVAILABLE

Utility Info. Avail	
Cross-References	
Comments	
File Number	PAG-NRC-038
Bib. Entry	Page, E. B., & Stake, R. E. (1979). Should educational evaluation be more objective or more subjective? *Educational Evaluation and Policy Analysis*, *1*, 45-47.
Confidentiality	
Eval. Design	A counterpoint article defending the objective vs subjective debate in evaluation is present. Page argues evaluation should be more objective to counteract the four weaknesses in the field: ethical dilemmas, measurement problems, training, and value laden technology. Stake argues for more subjective evaluation as it is *the* essential aspect of evaluation. A program's worth cannot be defined on achievement outcomes alone.
Eval. Type	
Eval. Model	
Evaluator Type	
Data Gen./Analysis	
Data Gath. Methods	
Data Analysis Tech.	
Utility Info.	
Intended Audiences	
Reporting Format	
Utility Info. Avail	
Cross-References	
Comments	
File Number	PAT-NRC-042
Bib. Entry	Patton, M. Q. (1980). Making Methods Choices. *Evaluation and Program Planning*, *3*, 219-228.
Confidentiality	
Eval. Design	The "paradigm" debate is discussed with emphasis on avoiding an either/or type of thinking when applying methodology in evaluation. The link between paradigm and methodology is questioned. Methods choices should be made based on evaluation in question.
Eval. Type	
Eval. Model	
Evaluator Type	
Data Gen./Analysis	
Data Gath. Methods	
Data Analysis Tech.	

BEST COPY AVAILABLE

Utility Info.	
Intended Audiences	
Reporting Format	
Utility Info. Avail	
Cross-References	
Comments	
File Number	RAY-NRC-046
Bib. Entry	Rayder, N. F. (1979). *Public outcry for humane evaluation and isomorphic validity*. Draft. San Francisco: Far West Laboratory for Educational Research and Development. (ERIC Document Reproduction Service No. ED187 710).
Confidentiality	
Eval. Design	Evaluation needs to be more humanistic especially to be "isomorphically" valid. Quotations from parents of school children are presented to support this statement. Twelve children are presented to promote humanistic evaluation: users should be involved; methods should be clear; individuals should be protected; evaluation should encourage use of information; information should be used for self-evaluation; on-going decisions for program improvement should be made; evaluation should document program responsiveness to the learner; evaluation should document treatment of individuals; evaluation should be designed to view learner developmentally; evaluation should include assessment of students and teachers; a human input statement should be included in the report; and the method should be congruent with service delivery. Three models of isomorphic validity are discussed.
Eval. Type	
Eval. Model	
Evaluator Type	
Data Gen./Analysis	
Data Gath. Methods	
Data Analysis Tech.	
Utility Info.	
Intended Audiences	
Reporting Format	
Utility Info. Avail	
Cross-References	
Comments	
File Number	SAD-NRC-070
Bib. Entry	Sadler, D. R. (1981). Intuitive data processing as a potential source of bias in naturalistic evaluations. *Educational Evaluation and Policy Analysis*, *3*, 425-31.
Confidentiality	

BEST COPY AVAILABLE

Eval. Design

Bias as a source of threat to an evaluation's validity is examined. Potential sources of bias include: 1) conflict of interest; 2) reactivity between evaluator and user; and 3) poor handling of evaluation. One way to deal with bias is via naturalistic inquiry. It is important there are limitations in terms of information-processing. These include: data overload, the effect of first impression, information availability, positive and negative instances, internal consistency, redundancy, novelty of information, reliability, missing data, revision of evaluation, proportion of population which findings describe, sampling, judgment confidence and consistency, and co-occurence. Knowing these limitations can lead to more effective evaluations using the naturalistic model.

Eval. Type
Eval. Model
Evaluator Type
Data Gen./Analysis
Data Gath. Methods
Data Analysis Tech.
Utility Info.
Intended Audiences
Reporting Format
Utility Info. Avail
Cross-References
Comments

File Number STR-NRC-080
Bib. Entry Strasser, S., & Deniston, O. L. (1978). Pre- and Post-planned evaluation: Which is preferable? Evaluation and Program Planning, 1, 195-202.

Confidentiality

Eval. Design

The authors compare pre- and post-planned evaluation approaches. These methods are compared across these dimensions: reliability; cost of collecting data; validity; evaluation obtrusiveness; and program goal displacement and direction. A model is presented to help program managers decide which model to use based on 3 decision questions: what will be the nature of the program when operationalized? What resources will be available? And, what steps are necessary to generate convincing findings?

Eval. Type
Eval. Model
Evaluator Type
Data Gen./Analysis
Data Gath. Methods

BEST COPY AVAILABLE

Data Analysis Tech.	
Utility Info.	
Intended Audiences	
Reporting Format	
Utility Info. Avail	
Cross-References	
Comments	

File Number	ZIM-NRC-001
Bib. Entry	Zimmerman, E. (1991). Authentic evaluation of progress and achievement of artistically talented students from diverse backgrounds. Unpublished manuscript, Indiana University.
Confidentiality	
Eval. Design	FO
Eval. Type	IN
Eval. Model	PC
Evaluator Type	
Data Gen./Analysis	
Data Gath. Methods	DA
Data Analysis Tech.	OL; OT

Utility Info.	Art learning is best assessed via "authentic means" conducted by art teachers "who are creators and consumers of assessment practices." Authentic assessment "attends to realistic situations of making and responding to works of art." Archbald and Newmann (1988) and Wiggins (1989) have established criteria for authentic assessment which can be applied to art: 1) evaluate tasks that approximate disciplined inquiry; 2) consider knowledge wholistically; 3) value achievement separate from assessment; 4) attend to process and products; 5) teach self-evaluation; 6) expect students to present and defend work; and 7) assess cooperation. Successful authentic measures of art include: exhibitions and performances. Gardner (1990) advocates "process portfolios", which are collections of student work both as final products and those in process in which students are involved with

BEST COPY AVAILABLE

	selection process. Process portfolios also allows for assessment of risk-taking, problem-solving, and evaluation of self and others. Another form of authentic assessment includes the use of profiles of behaviors to assess work habits, learning abilities, knowledge, skills, and interest. Journal entries and interviews provide other means. The assessment techniques previously mentioned are also discussed in light of how to assess students from diverse backgrounds.
Intended Audiences	TE; ST; PA
Reporting Format	OT
Utility Info. Avail	
Cross-References	
Comments	

BEST COPY AVAILABLE

Appendix F

A Planning Guide for Evaluating Programs for Gifted Learners

Quest

Published by the Division of Research and Evaluation
A Division of the National Association for Gifted Children

The Division of Research and Evaluation
Volume 4, Number 2
Fall 1993

Message From the Chair

Rena F. Subotnik
Hunter College

This is my last column as Chair of Research and Evaluation. The leadership of the Division will be in the able hands of Bonnie Cramond after the 1993 convention. I hope to model after past-chair, Paula Olszewski, by remaining involved with Divisional activities.

This issue includes a nomination form(page 15). There are two very important slots for which we have no nominees. Each year we elect an Assistant Program Chair. This position is designed to prepare the holder to smoothly take over the position of Program Chair for the 1995 convention. We are one of the largest and most active divisions; therefore, the logistics of reviewing and selecting proposals and organizing our traditional events is one we all value enormously. Famous past program chairs include Gina Schack, Alane Starko, and Marcie Delcourt. I'm sure you are aware of the fine work of our current Program Chair, Richard Olenchak, ably assisted by Sherry Wilson.

The other position is that of Chair-Elect. We need a nominee to proudly represent the organization in the Division Steering Committee, an important forum for divisional issues, and to guide the group discussion at our business meeting. Please consider this position if you have strong attachments to the Division and to our role in NAGC.

Lynne Hannah, Secretary, and Sidney Moon, Newsletter Editor, have volunteered to run again. I urge you to support the Division by voting for them at the election or to nominate other members who have the skills and desire to take on these roles.

I hope that you have been refreshed by your summer activities and that this coming year is productive, healthy, rewarding for you. See you in Atlanta.

A Planning Guide for Evaluating Programs for Gifted Learners

Carol A. Tomlinson
Carolyn M. Callahan
The University of Virginia

The work reported herein was sponsored by the National Research Center on the Gifted and Talented under the Jacob K. Javits Gifted and Talented Students Education Act (Grant No. R206R00001) and administered by the Office of Educational Research and Improvement and the United States Department of Education. The findings do not reflect the positions or policies of the Office of Educational Research and Improvement or the United States Department of Education.

Educational accountability is a popular topic in political circles, but in practice effective evaluation of school programs is sporadic at best. The field of gifted education appears especially problematic in this regard. There appear to be relatively few examples of robust evaluation designs and procedures currently in use with programs for the gifted (Hunsaker and Callahan, 1993). Among reasons for the paucity of effective evaluation practices in programs for the gifted are weakness of evaluation skill among directors of such programs, lack of time and funding required for meaningful evaluation, complex problems posed in appropriately evaluating the kinds of learning outcomes typical of programs for the gifted, and fear of public discussion of programming for gifted learners where funding for gifted education is tenuous (Tomlinson, Bland, Moon and Callahan, 1992). The evaluation literature is full of recommendations, models and admonitions about appropriate practice. Indeed the literature can easily overwhelm anyone trying to decide on the most

Focus on Evaluation

In this issue...

A Planning Guide for Program Evaluation........1
Book Corner........5
Keeping Up to Date........6
Research Starters........9
Fellowships........14
Nomination Form........15

BEST COPY AVAILABLE

fundamental concerns in designing a useful and valid evaluation plan. The purpose of this article is to draw together insights from research in general educational evaluation and evaluation of programs for the gifted (e.g., Joint Committee on Standards for Evaluation, 1981; Tomlinson, C., Bland, L., & Moon, T., 1993) in order to provide systematic guidance based on the most effective practice for those charged with planning, executing, and using findings from evaluations of educational programs for gifted learners. Evaluation should proceed through four stages; (1) preparing for the evaluation, (2) designing data collection and analysis, (3) conducting the evaluation, and (4) reporting findings and follow-up. The questions presented here provide a guide based on research and effective practice to aid those planning evaluation for programs for the gifted. The framework provided in the planning guide should, of course, be modified to address specific evaluation needs in a given setting.

A Planning Guide For Evaluating Programs for Gifted Learners

Carol A. Tomlinson
Carolyn M. Callahan
National Research Center on the Gifted and Talented
The University of Virginia

(This guide is based on research and best practices both in the field of general educational evaluation and evaluation of programs for the gifted. It poses questions intended to facilitate the thinking and planning of individuals and groups charged with evaluating programs for gifted learners. Those using the guide are encouraged to modify it in ways which make the evaluation process better tailored to address local needs and concerns.)

Preparing for the Evaluation

(Much of the success of a program evaluation will depend on the quality of decisions made prior to actually conducting the evaluation. Planning is an essential phase of the process and should proceed carefully and thoughtfully.)

- Does the program have clearly articulated goals and objectives which can be a focus of evaluation?
- Are the articulated goals and objectives the ones valued as a program focus?
- Does the school division have a commitment to meaningful evaluation of programs including adequate time, finances and personnel time given to evaluation and dissemination of findings?
- Have you identified representatives of varied internal and external interest groups or stakeholders (i.e. parents, regular classroom teachers, administrators, students, gifted/talented specialists, school board members, representatives of business and industry, etc.) to serve as an active evaluation steering committee which will be involved in setting the parameters of the evaluation?
- Is there a written plan for evaluating the program, including delineated steps and procedures in the process?
- Is there a plan for on-going feedback during the evaluation (formative as well as summative evaluation)?
- Are the evaluators knowledgeable about both gifted education and evaluation?
- Are the evaluators knowledgeable about both qualitative and quantitative research strategies?
- Do evaluators, program personnel and/or steering committee members include those with sufficient political sophistication to understand the political implications of evaluation? Can they aid in identifying and gaining access to key decision makers and can they provide an understanding of the actions over which the decision-makers have control?
- Are roles of evaluators, administrators, stakeholders and steering committee members in the evaluation process clearly articulated?
- Is there a working plan to develop networks of support both inside and outside the school division for the evaluation process, its findings, and the program?
- Are there appropriate timelines for data gathering, analysis and dissemination?
- Will the evaluation data be collected, analyzed and presented in time to influence decision-making?
- Are there plans and procedures for monitoring processes and procedures throughout the evaluation?
- Are appropriate provisions established to ensure confidentiality and sensitivity in handling data?

Designing Data Collection

(Designing evaluations for programs for gifted learners is difficult because of the complex nature of instructional interventions appropriate for gifted learners and the shortcomings of traditional standardized measures in reflecting the impact of such interventions. It is important for evaluators of programs for gifted learners to carefully match evaluation goals with data collection modes capable of demonstrating student growth.)

- Are there clearly stated evaluation questions which clearly and appropriately address program goals, structures, functions, and/or activities?
- Do the evaluation questions seem likely to generate findings which will have a positive impact on programs and participants?
- Are there plans to use multiple data sources (e.g. parents, regular classroom teachers, identified students, other students, gifted education specialists, administrators) in order to understand perspectives of various stakeholders?
- Are there plans to employ

BEST COPY AVAILABLE

varied data collection modes (e.g. face-to-face interviews, telephone interviews, classroom observations, group meetings, product reviews, staff development evaluations, mail out surveys, test data, etc.) in order to reflect the complex nature of the program and meet data needs of various constituencies?

- Do potential users of findings have opportunities to provide input on types of information desired and forms in which the information would be most usefully reported?
- Have you examined ways to collect "process data" which can show whether the program is functioning as it should?
  - attendance records
  - documents (agendas, minutes, handouts, etc.) from g/t staff meetings, Parent Advisory Committee meetings, division-wide staff meetings
  - communications between school and home
  - communications between g/t program and regular program
  - observation data from g/t class collected by qualified observers documenting what takes place in the program or curricular modification being studied
  - teacher and/or student journals
  - g/t teacher lesson plans or other planning documents
  - description of regular class and special class settings in regard to gifted learners via checklist utilized by qualified observer
  - attitude data (e.g. interviews, surveys, etc.) which allow various stakeholders to indicate their perceptions of the program's effectiveness
- Have you examined ways to collect "outcome data" which can show whether student affective and/or academic growth has occurred as a result of program participation?
  - comparison between aptitude and achievement measures of eligible program participants and eligible program non-participants
  - use of out-of-level achievement data with program participants
  - use of comparison groups (including varying times when participating students receive interventions so that, for example, students in one g/t class receive an intervention first semester and those in another serve as a control group first semester and receive the intervention second semester)
  - portfolio/product rating according to predetermined criteria by experts with demonstrated inter-rater reliability
  - use of "retrospective pretesting" in which program participants reflect on specific ways in which their knowledge and skill have changed as a result of program participation
  - traditional experimental or quasi-experimental designs with control and treatment groups (including evidence of achievement of identified students when the same topics are explored through regular class and special class settings)
  - use of valid and reliable self-concept inventories with control and treatment groups and/or as pre and post data for a single group
- Have you considered ways in which case study data can be useful to document program effectiveness?
- Have you selected reliable and valid assessment tools?
- Have you described ways in which data will be analyzed?
- Have you specified ways in which data will be reported to various groups?
- Have you prepared staff members for the data-collection phase of the evaluation process and their roles in it?

Conducting the Evaluation

(While the evaluation is being conducted, there is a great need for continued involvement of evaluators and the steering committee to ensure appropriate management of data and use of findings, and to ensure involvement of appropriate groups and individuals in the process.)

- Are multiple stakeholders consistently involved with data collection?
- Are program evaluators consistently visible to varied audiences to facilitate understanding of those audiences by the evaluators and understanding of the program and evaluation process by the audiences?
- Are multiple stakeholders consistently involved with monitoring and reviewing the evaluation process and its evolving findings?
- Do you have a plan for quick turnaround time for data analysis and feedback, with specific guidelines for all individuals in meeting prescribed timelines?
- Is there a commitment from evaluators, key program personnel and steering committee members to use of findings for positive program change?
- Is there an articulated plan for turning findings into action, incorporating the roles which evaluators, program personnel and stakeholders will play in that process?

Reporting Findings and Follow-Up

(Evaluations are useful only if their findings result in positive change for programs and participants. Findings must be made available in appropriate forms to varied stakeholder groups and plans of political action must be developed and followed.)

- Have evaluators, program personnel and evaluators assessed the impact of evaluation findings?
- Are findings prepared and interpreted according to interest and needs of stakeholder groups?

Continued on page 4

- Are evaluation reports clear? Do they avoid use of jargon and confusing, technical interpretations of data?
- Do evaluation reports describe the program, evaluation questions, evaluation process, participants in the process, data collection, and data analysis?
- Are evaluation reports designed for follow-through with specific recommendations made for acting upon findings?
- Are evaluation reports and recommendations presented to decision-makers in a timely fashion?
- Are there provisions for oral explanations and discussions of findings with stakeholders and decision-makers?
- Has the steering committee assessed the evaluation process according to initial goals, roles and timelines, including making written recommendation for changes in the next evaluation cycle?
- Have evaluators, steering committee members and program personnel followed up with policy makers until appropriate actions have been taken?
- Has the steering committee proposed questions for further examination in upcoming evaluation cycles and resulting from insights gained in the current evaluation cycle?

References

Hunsaker, S., & Callahan, C. (1993). Evaluation of gifted programs: Current practices. *Journal for the Education of the Gifted; 16*, 190-200.

Joint Committee on Standards for Educational Evaluation. (1981). *Standards for evaluations of educational programs, projects, and materials*. New York: McGraw-Hill.

Tomlinson, C., Bland, L., & Moon, T. (1993). Evaluation utilization: A review of the literature with implications for gifted education. *Journal for the Education of the Gifted, 16*, 171-189.

Tomlinson, C., Bland, L., Moon, T., & Callahan, C. (1992). Evaluation designs and practices: Case studies in gifted education. Manuscript submitted for publication.

1993 Research Briefs

will be available

at NAGC!

OTHER RESEARCH BRIEFS STILL AVAILABLE!

ORDER YOUR COPIES NOW!

1992 Research Briefs $9.00*

Make your check payable to NAGC (don't forget shipping and handling charges) and mail your order to:

National Association for Gifted Children
1155 15th Street, NW
Suite 1002
Washington, D.C. 20005

*Shipping and handling charges:

$0-$9.99	$2.25
$10-$24.99	$3.25
$25-$49.99	$4.25
$50-$74.99	$5.25
$75-$99.99	$6.25
$100-$999.99	7% of net order

(GREAT FOR GRADUATE SEMINARS)

Quest **Volume 4, Number 2**

A publication of the Research and Evaluation Division
National Association for Gifted Children
Published bi-annually
Sidney M. Moon, Editor, Purdue University
Department of Educational Studies, 1446 LAEB
West Lafayette, IN 47907-1446
FAX: 317-494-1228

Design and Layout ---------------------- Susan Carr
Editorial Assistant ---------------------- Madonna Starkey
Column Editors ------------------------- Susan Linnemeyer
-- Karen Rogers
-- Micheal Sayler
Reviews Editor ------------------------- Susan Assouline

BEST COPY AVAILABLE

Appendix G

Guidelines for Conducting Useful Evaluations of Programs for Gifted Learners

GUIDELINES
for Conducting *Useful* *Evaluations of Programs* for Gifted Learners

Prepared by
National Research Center
on the Gifted & Talented
The University of Virginia

1. **Make evaluation a part of planning from the earliest stages of program development.**

   A program which builds in systematic processes and timelines for evaluation is more likely to yield data which are useful to varied stakeholders and aimed toward positive program change.

   A commitment to evaluation by the gifted education staff is essential but needs also to be accompanied by a clear division-wide expectation that all program areas will be evaluated regularly and appropriately. Without a commitment on the part of people in positions of power and influence that evaluation should result in program change, little attention is likely to be given to evaluation findings.

2. **Develop clear program descriptions and goals.**

   These should provide a road map for evaluation as you seek to determine whether the program is meeting specific goals and is functioning as it is described. Be sure goals are specific, focused and clear, and that descriptions are accurate.

3. **Provide adequate funding for evaluations and adequate time for evaluation procedures to be followed.**

   It is unlikely that a broadly useful evaluation will be conducted in the absence of funding for preparation of evaluation materials,. support personnel, data processing, etc. Also a well-planned evaluation will require ample time in order to involve key stakeholders and to assess varied aspects of program function.

4. **Prepare staff for conducting and analyzing the results of the evaluation.**

   In evaluating programs for the gifted, it is important that persons knowledgeable of both evaluation and gifted education play lead roles throughout the evaluation. It is likely in many school divisions that key personnel will need meaningful training in one or both areas.

5. **Clearly identify all audiences who have an interest in or need for evaluation results, and involve them in the full evaluation process.**

   Involvement of multiple stakeholders throughout the process gives more people a sense of ownership of both the program and its outcomes, and yields more advocates for positive program change stemming from evaluation findings. Be sure to include relevant policy makers in the group of stakeholders.

6. **Ask questions which are well focused to provide information about the goals, structures, and activities of the program being evaluated–questions which will aid in making significant program improvements.**

BEST COPY AVAILABLE

7. Use multiple data sources in order to understand the values and perspectives of varied groups of stakeholders.

   Members of school boards, building and central office administrators, identified students, regular classroom teachers, g/t program staff, counselors, and many others will be able to give interesting insights into program functioning

8. Develop evaluation designs which address complex issues of measurement in programs for the gifted.

   Assess both process outcomes (are components of the program functioning as they should) and product outcomes (are students growing academically and/or affectively as a result of the program).

   Quantitative designs may be more effective in looking at product outcomes and qualitative designs may be more effective in looking at process outcomes.

   Avoid reliance on traditional standardized measures which offer little promise of reflecting academic growth in gifted learners (consider instead options such as retrospective pretesting, use of contrast groups or carefully matched groups, out-of-level testing, a time-series design, use of interventions in both regular classes and g/t classes to measure the breadth and depth of achievement and rate of learning in the two classes, etc.

9. Use a variety of data gathering methods designed to reflect the unique structure and goals of programs for gifted learners (e.g. out-of-level testing, portfolio assessment, product rating with common criteria and demonstrated inter-rater reliability, qualitative studies which describe unique settings, surveys, observation checklists, etc.)

10. In evaluation reports, describe fully procedures for data collection and interpretation so that audiences understand processes which were followed and conclusions which were drawn.

11. Disseminate to all appropriate audiences reports which are timely and designed to encourage follow-through in translating findings into action. Develop a specific plan for turning findings into positive program growth as an essential part of each evaluation, including roles which various program personnel, evaluators and stakeholders will play in that plan.

BEST COPY AVAILABLE

For additional information, you may request one of the following papers from the National Research Center on the Gifted and Talented at the University of Virginia:

Tomlinson, C., Bland, L., & Moon, T. (in press). Evaluation utilization: A review of the literature with implications for gifted education.

Tomlinson, C., Bland, L., Moon, T., & Callahan, C. (1992). Designing user-friendly evaluations for programs for the gifted. Manuscript submitted for publication.

Tomlinson, C., Bland, L., Moon, T., & Callahan, C. (1992). Evaluation designs and practices: Case studies in gifted education. Manuscript submitted for publication

NRC
G/T

For additional information on research projects leading to these recommendations, contact:
Carol Tomlinson, Ed.D.
National Research Center on the Gifted and Talented
Room 287 Ruffner Hall
The University of Virginia
Charlottesville, VA 22903

The work reported herein was sponsored by the National Research Center on the Gifted and Talented, sponsored under the Jacob K. Javits Gifted and Talented Students Education Act (Grant No. R206R00001) and administered by the Office of Educational Research and Improvement and the United States Department of Education. These findings do not reflect the positions or policies of the Office of Educational Research and Improvement or the United States Department of Education.)

BEST COPY AVAILABLE

Appendix H

Reference on Evaluation Utility Classified According to the Factors Established by the Joint Committee on the Standards for Educational Evaluation[*]

---

[*] Joint Committee on the Standards for Educational Evaluation. (1981). *Standards for evaluations of educational programs, projects, and materials.* New York: McGraw-Hill.

Standard. Audience Identification: Audiences involved in or affected by the evaluation should be identified, so that their needs can be addressed.

References. Alkin (1980); Ball and Anderson (1977); Bissell (1979); Buescher (1986); Caulley (1981); Cox (1977); Eichenberger (1979); Fleischer (1984); Franchak and Kean (1981); Kilburg (1980); Leviton and Hughes (1981); Marshall (1984); Mathis (1980); Raizen and Rossi (1981); Wolf (1980).

Standard. Evaluator Credibility: The persons conducting the evaluation should be both trustworthy and competent to perform the evaluation, so that their findings achieve maximum credibility and acceptance.

References. Alkin and Ruskus (1981); Ball and Anderson (1977); Franchak and Kean (1981); Kingsbury (1980); Leviton and Hughes (1981); Marshall (1984); Patton (1988); Stalford (1979).

Standard. Information Scope and Sequence: Information collected should be of such scope and selected in such ways as to address pertinent questions about the object of the evaluation and be responsive to the needs and interests of specified audiences.

References. Alkin and Ruskus (1981); Apling (1981); Caulley (1981); Cox (1977); Franchak and Kean (1981); Kingsbury (1980); Leviton and Hughes (1981); Marshall (1984); Nguyen (1978); Raizen and Rossi (1981).

Standard. Valuational Interpretation: The perspectives, procedures, and rationale used to interpret the findings should be carefully described, so that the bases for value judgments are clear.

References. Alkin and Ruskus (1981); Ball and Anderson (1977); Caulley (1981); Cox (1977); Deniston (1980); Englert, Kean, and Scribner, (1977); Franchak and Kean (1981); Leviton and Hughes (1981); Marshall (1984); Raizen and Rossi (1981); Smith (1981).

Standard. Report Clarity: The evaluation report should describe the object being evaluated and its context, and the purposes, procedures, and findings of the evaluation, so that the audiences will readily understand what was done, what information was obtained, what conclusions were drawn, and what recommendations were made.

References. Alkin and Ruskus (1981); Apling (1981); Ball and Anderson (1977); Caulley (1981); Cox (1977); Eichenberger (1979); Franchak and Kean (1981); Kingsbury (1980); Leviton and Hughes (1981); Marshall (1984); Nguyen (1978); Raizen and Rossi (1981).

Standard. Report Dissemination: Evaluation findings should be disseminated to clients and other right to know audiences, so that they can assess and use the findings.

References. Ball and Anderson (1977); Caulley (1981); Cox (1977); Dickey and Hampton (1981); Englert, Kean, and Scribner, (1977); Franchak and Kean (1981); Marshall (1984); Raizen and Rossi (1981).

Standard. Report Timeliness: Release of reports should be timely, so that audiences can best use the reported information.

References. Alkin and Ruskus (1981); Ball and Anderson (1977); Caulley (1981); Cox (1977); Englert, Kean, and Scribner, (1977); Franchak and Kean (1981); Kingsbury (1980); Marshall (1984); Nguyen (1978); Raizen and Rossi (1981); Smith (1981).

Standard. Evaluation Impact: Evaluations should be planned and conducted in ways that encourage follow-through by members of audiences.

References. Alkin and Ruskus (1981); Apling (1981); Ball and Anderson (1977); Bissell (1979); Brown, Newman, and Rivers (1984); Caulley (1981); Cox (1977); Fleischer (1984); Franchak and Kean (1981); Leviton and Hughes (1981); Marshall (1984); Patton (1988); Raizen and Rossi (1981); Smith (1981); Stalford (1979); Wolf (1980).

Appendix I

Second Round of Interview Questions

Interview Questions

We are aware that within the last couple of years an evaluation of the gifted program in your school district was conducted, please tell us about the process of this evaluation and its outcome?

How did the evaluation affect your thinking about the program?

How was the evaluation information used?

Additional Questions

How did the evaluation influence program development positively?

How did the evaluation influence program development negatively?

What other influences did the evaluation have for program development?

Possible Factors to Consider

Was the evaluation timely in reference to making a difference for the budget?

How quickly was the evaluation done?

What was the background and training of the evaluator?

What types of evaluation had the evaluator done previously?

Were examples of implementation for program change included?

How much money did the evaluation cost?

How much money did the recommended changes cost?

Did people perceive that too much money was already being spent on the program?

Was enough money spent to pay attention and believe the evaluator?

Were the evaluators knowledgeable about the field?

Were the evaluators knowledgeable of changes in the program?

What are the benefits for students if change is made?

Did the recommendations demand additional teacher time?

Was the report readable?

Did the evaluator know enough to keep away from scared cows?

Did the evaluator become a stakeholder?

What particular model of gifted education did the evaluator buy into?

Were the results communicated to parents, teachers, the community?

Were the recommendations made to change negative aspects of the program into positive aspects?

Were the recommendations based on research?

Appendix J

Evaluation Instruments Database Form

THE NATIONAL RESEARCH CENTER ON THE GIFTED AND TALENTED
DATA BANK SEARCH REQUEST FORM

BIBLIOGRAPHIC INFORMATION: EVALUATION

This form is to be used for requesting annotated bibliographies on procedures and tests used in the evaluation of programs for gifted and talented students. Because our list is so extensive we ask you to specify the types of information you are looking for by completing this form. If you are seeking a review of a specific tests, please use the form labelled **"TEST REVIEW REQUEST: EVALUATION"**, if you are seeking a list of such tests, please use the **"TEST INFORMATION REQUEST: EVALUATION"**

I. State the type of information you are seeking by filling in the cost next to each item for which you want information:

_______ Information on instrumentation ($7.50)
_______ Information on evaluation designs ($7.50)
_______ Information on evaluation issues ($7.50)
_______ Information on evaluation utility ($7.50)
_______ Information on needs assessments ($7.50)

======= **Page Total** (Please transfer to Order Summary Page and return both pages.)

II: State the objective or goal you are seeking to measure in the evaluation process. These may range from student outcome goals (e.g., Students are more independent as a result of involvement in the Quest program) to process goals (e.g., Teachers engage students in higher level thinking processes), to management goals (e.g., Parents are well-informed about the curriculum of the program). Please state no more than one goal per request. Use a separate order form for each goal for which you want information.

BEST COPY AVAILABLE

The sections below allow searches to be refined to better meet your needs. If you indicate specific areas of interest here, your search will be limited to instruments used in these specific ways.

II. Grade level

______ Preschool
______ Primary (K-2)
______ Elementary (K-5)
______ Middle school (6-8)
______ High school (9-12)

III. Specific target population

______ African-American/Black
______ Hispanic/Latino
______ Native American/American Indian
______ Asian-American
______ Polynesian
______ Handicapped/Learning disabled
______ Handicapped/Hearing impaired
______ Handicapped/Visually impaired
______ Handicapped/Physically challenged
______ Other (please specify:____________________)

BEST COPY AVAILABLE

THE NATIONAL RESEARCH CENTER ON THE GIFTED AND TALENTED
DATA BANK SEARCH REQUEST FORM

TEST INFORMATION REQUEST: EVALUATION

This form is to be used for requesting a list of tests used in the evaluation of programs for the gifted and talented. Because our list is so extensive, we ask you to specify the types of tests you are looking for by completing this form. If you are seeking a review of a specific test, please use the form labelled "**TEST REVIEW REQUEST: EVALUATION**".

This list will contain all instruments that have been reported as used for the purpose stated. Evaluations of the instruments are not included in this list. If you wish specific evaluations of specific tests after receiving the list, you may request that information from us.

I: Indicate the objective(s) or goal(s) you are seeking to measure in the evaluation process. These may range from student outcome goals (e.g., Students are more independent as a result of involvement in the Quest program) to process goals (e.g., Teachers engage students in higher level thinking processes), to management goals (e.g., Parents are well-informed about the curriculum of the program). The cost is $7.50 per goal/objective assessed.

_______ 1.

_______ 2.

_______ 3.

_______ 4.

======= **Page Total** (Please transfer to Order Summary Page and return both pages.)

BEST COPY AVAILABLE

The sections below allow searches to be refined to better meet your needs. If you indicate specific areas of interest here, your search will be limited to instruments used in these specific ways.

II. Grade level:

______ Preschool
______ Primary (K-2)
______ Elementary (K-5)
______ Middle school (6-8)
______ High school (9-12)

III. Specific target population

______ African-American/Black
______ Hispanic/Latino
______ Native American/American Indian
______ Asian-American
______ Polynesian
______ Handicapped/Learning disabled
______ Handicapped/Hearing impaired
______ Handicapped/Visually impaired
______ Handicapped/Physically challenged
______ Other (please specify:________________)

IV. Type of instrument

______ Standardized, objective test
______ Locally developed objective test
______ Rating scale or checklist
______ Portfolio
______ Other (please specify: ________________)

V. Expected respondent (Whom do you wish to gather information from?) (please check all that apply):

_____ Students
_____ Parents
_____ Teachers of the gifted
_____ Administrators
_____ School Board Members
_____ Regular classroom teachers
_____ Other

BEST COPY AVAILABLE

THE NATIONAL RESEARCH CENTER ON THE GIFTED AND TALENTED
DATA BANK SEARCH REQUEST FORM

TEST REVIEW REQUEST: EVALUATION

This form is to be used for requesting reviews of specific tests used in the evaluation of programs for gifted and talented students. If you are seeking a listing of tests used in evaluating gifted programs, please use the form labelled "TEST INFORMATION FORM: EVALUATION".

Complete Name of the Test: ______________________________
Publisher: (if known): ______________________________
Form (if applicable): ______________________________

Goal(s) or objective(s) of the program that you are seeking to assess. The cost for each goal/objective assessed is $7.50.

______ 1.

______ 2.

______ 3.

______ 4.

======= **Page Total** (Please transfer to Order Summary Page and return both pages.)

BEST COPY AVAILABLE

THE NATIONAL RESEARCH CENTER ON THE GIFTED AND TALENTED
DATA BANK SEARCH REQUEST FORM

LOCAL INSTRUMENT REQUEST: EVALUATION

This form is to be used when requesting copies of instruments developed by individual school divisions for use in their own evaluation process. These school divisions have generously allowed their materials to be shared through the NRC/GT Data Banks. If you wish lists of standardized instruments used by schools, please use the form "**TEST INFORMATION REQUEST: EVALUATION.**"

In order to provide you with the most helpful information, our collection of instruments is divided according to the area of giftedness the program emphasizes. These divisions are further categorized according to various aspects of the evaluation process (eg, formative or summative evaluation or the instrument respondent). Instruments are available in sets of three for $3.00 or a set of six for $6.00. For some areas of giftedness, we may not be able to provide six instruments: these are marked "**THREE ONLY**" on the list below.

I: Select the area of giftedness, category of giftedness, or attribute you are emphasizing in the evaluation process. In the line to the left of the attribute, write $3.00 if you wish three instruments or $6.00 if you wish 6 instruments. If you wish to limit your search to specific grade levels, special populations, or respondent, be sure to indicate your choice(s) on the next page.

_______ verbal/linguistic achievement
_______ mathematical/logical achievement
_______ scientific achievement
_______ social sciences achievement
_______ visual arts ability (Please specify:______________________)
_______ performing arts ability (please specify:______________________)
_______ vocational education/practical arts ability -- **THREE ONLY**
_______ self-concept/self-esteem -- **THREE ONLY**
_______ attitude towards school -- **THREE ONLY**
_______ creativity: ideation
_______ creativity: problem-solving
_______ task commitment/motivation -- **THREE ONLY**
_______ critical thinking -- **THREE ONLY**

======= **Page Total** (Please transfer to Order Summary Page and return both pages.)

BEST COPY AVAILABLE

The sections below allow searches to be refined to better meet your needs. If you indicate specific areas of interest here, your search will be limited to instruments used in these specific ways.

II. Grade level

_______ Preschool
_______ Primary (K-2)
_______ Elementary (K-5)
_______ Middle school (6-8)
_______ High school (9-12)

III. Respondent

______ Teacher
______ Parent
______ Student/Peer
______ Administrator
______ School Psychologist
______ Community Leader
______ Other (Please specify: ______________________________)

IV. Evaluation Type

______ Formative
______ Summative
______ Needs Assessment

V. Program Type

______ Pullout
______ Within Class
______ Special Class
______ Special School
______ After School/Saturday/Summer

VI. Program Aspect

______ Specific Subject Area Content Knowledge
______ Process Skills
______ Student Products
______ Social and/or Affective Effects

BEST COPY AVAILABLE

Research Monograph Series

The National Research Center on the Gifted and Talented
The University of Connecticut
362 Fairfield Road, U-7
Storrs, CT 06269-2007

Production Assistants
Dawn R. Guenther
Siamak Vahidi

Reviewers
John F. Feldhusen
Reva Friedman
E. Jean Gubbins
Mary G. Rizza
Michael J. Spottiswoode
James T. Webb

Also of Interest

Regular Classroom Practices With Gifted Students: Results of a National Survey of Classroom Teachers
Francis X. Archambault, Jr., et al.

An Observational Study of Instructional and Curricular Practices Used With Gifted and Talented Students in Regular Classrooms
Karen L. Westberg, et al.

Why Not Let High Ability Students Start School in January?
The Curriculum Compacting Study
Sally M. Reis, et al.

Evaluation of the Effects of Programming Arrangements on Student Learning Outcomes
Marcia A. B. Delcourt, et al.

Qualitative Extension of the Learning Outcomes Study
Marcia A. B. Delcourt and Karen Evans

Also of interest from the

Research Monograph Series

Toward a New Paradigm for Identifying Talent Potentials
Mary M. Frasier and A. Harry Passow

Talents in Two Places: Case Studies of High Ability Students With Learning Disabilities Who Have Achieved
Sally M. Reis, Terry W. Neu, and Joan M. McGuire

The Effects of Group Composition on Gifted and Non-Gifted Elementary Student in Cooperative Learning Groups
David A. Kenny, Francis X. Archambault, Jr., and Bryan W. Hallmark

Extending the Pedagogy of Gifted Education to All Students
Sally M. Reis, Marcia Gentry, and Sunghee Park

Case Studies of Talented Students Who Achieve and Underachieve in an Urban High School
Sally M. Reis, et al.

Profiles of Successful Practices for High Ability Students in Elementary Classrooms
Karen L. Westberg and Francis X. Archambault, Jr.

Academic Diversity in the Middle School: Results of a National Survey of Middle School Administrators and Teachers
Tonya Moon, Carol A. Tomlinson, and Carolyn M. Callahan

A Triarchic Approach to Giftedness
Robert J. Sternberg

A Study of Achievement and Underachievement Among Gifted, Potentially Gifted, and Average African-American Students
Donna Y. Ford

Instruments Used in the Identification of Gifted and Talented Students
Carolyn M. Callahan, et al.

NRC
G/T

The National Research Center on the Gifted and Talented Research Teams

The University of Connecticut

Dr. Francis X. Archambault, Jr., Associate Director
The University of Connecticut
School of Education, U-4
Storrs, CT 06269-2004
860-486-4531

Dr. Alexinia Y. Baldwin
Dr. Scott W. Brown
Dr. Deborah E. Burns
Dr. David A. Kenny
Dr. Jonna Kulikowich
Dr. Sally M. Reis
Dr. Karen L. Westberg
Dr. Michael F. Young

The University of Georgia

Dr. Mary M. Frasier, Associate Director
The University of Georgia
Department of Educational Psychology
323 Aderhold Hall
Athens, GA 30602-7146
404-542-5106

Dr. Scott L. Hunsaker

The University of Virginia

Dr. Carolyn M. Callahan, Associate Director
Curry School of Education
The University of Virginia
405 Emmet Street
Charlottesville, VA 22903
804-982-2849

Dr. Michael S. Caldwell
Dr. Marcia A. B. Delcourt
Dr. Kathleen May
Dr. Claudia Sowa
Dr. Ellen Tomchin
Dr. Carol A. Tomlinson

Yale University

Dr. Robert J. Sternberg, Associate Director
Department of Psychology
Yale University
P.O. Box 208205
New Haven, CT 06520-8205
203-432-4632

Dr. Pamela Clinkenbeard